Urdu Text Extraction Method from Images

S. Tehsin, S. Kausar

***Abstract*—**Due to the vast increase in the multimedia data in recent years, efficient and robust retrieval techniques are needed to retrieve and index images/ videos. Text embedded in the images can serve as the strong retrieval tool for images. This is the reason that text extraction is an area of research with increasing attention. English text extraction is the focus of many researchers but very less work has been done on other languages like Urdu. This paper is focusing on Urdu text extraction from video frames. This paper presents a text detection feature set, which has the ability to deal up with most of the problems connected with the text extraction process. To test the validity of the method, it is tested on Urdu news dataset, which gives promising results.

***Keywords*—**Caption text, Content- Based Image Retrieval, Document analysis, Text extraction

I. Introduction

In recent years there is a fast increase in multimedia data, which instruct to improve the mechanisms to efficiently index and retrieve the huge amount of data. Video data retrieval and indexing is most difficult task, because of its huge and ever increasing amount of information. Video and image data is classically index and retrieve through its meta-data associated with; in the form of textual information. But that textual description cannot fully describe the contents of that image/video. So the best is to use the content of multimedia data for efficient retrieval systems. Textual information embedded in images and video can be the best content describer of particular multimedia data. Examples of such describer can be news credits, movie titles, product labels and sign board writings.

Text appearing in video/images can be classified in two groups naming Caption and Scene text. Caption text is computer generated superimposed text that is not actual part of that image or video. Where, scene text is actual part of the image, when captured. Text recognition process comprises of five phases. These phases are represented in figure 1.

Urdu is a language of the Indian subcontinent people. By an estimation, 130–270 million people speak Urdu. Writing script of Urdu is very much similar to Arabic, as both are written using Nastaliq calligraphic style of Persian-Arabic script.

A lot of multimedia stuff is available having the Urdu text embedded in it; including news, movies and TV commercials etc. This paper presents efficient and robust method for Urdu text extraction from images and videos.

S.Tehsin is with the CS Dept of Bahria University, Islamabad, pakistan (e-mail: tsamabia@yahoo.com).

S. Kausar is with the CS Dept of Bahria University, Islamabad, pakistan (e-mail: sum_satti@yahoo.com).

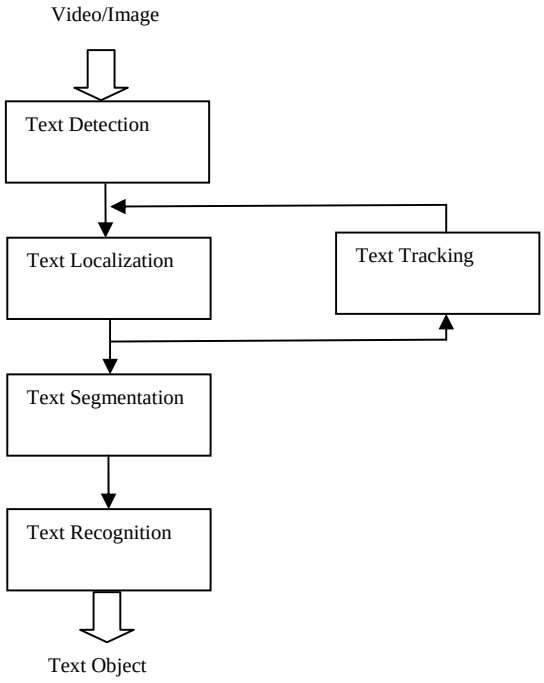

Figure 1: Architecture of text extraction and recognition process

II. Literature Review

A variety of approaches of text extraction have been proposed during the past years [1][6]. Comprehensive surveys can be found in [7][9]. On the basis of text feature utilization, these techniques can be mainly classified into two types: region based and texture based [10].

Texture-based techniques are based on the fact that text objects are having distinguishing textural features that differentiate it from the background. The techniques mostly use Gabor filters, Wavelet, FFT, spatial variance, etc. These methods further use machine learning techniques such as SVM, MLP and adaBoost [11][15].

Region based approach exploits different region properties to extract text objects. This approach makes use of the fact that there is sufficient difference between the text color and its immediate background. Color features, edge features, and connected component methods are often used in this approach [16][19].

A variety of work has been done on caption text extraction, but most of the work is application specific and tested on non standard datasets. Shuicai Shi et al [20] proposed an approach for text detection, localization and extraction in video frames using block change rate and element image division in CIE L*a*b*. This approach extracts multilingual text from videos.

Jiangbo Xu et al [21] proposed text extraction in DCT compressed domain. Candidate text blocks are detected in terms of DCT texture energy. An adaptive temporal constraint method is proposed to exploit the temporal occurrence of text in a sequence of frames. Results are verified on MPEG video sequences.

The merger of binarized and intensity images edge maps is used in Shi Jianyong [22] method to detect text in video frames. This method is time saving specially for monochromatic images. Xin Zhang [23] proposed text extraction method for multilingual text carrying images/video frames. In this method color and edge features are combined to extract the text. A two stage video-text location method is proposed by Wang Zhiming [24]. In the first stage, an unsupervised paradigm based on wavelet is proposed to obtain candidate text region. Text boundaries are marked in the second stage by traversing line with its aptitude spectrum.

Some application specific text detection approaches are also reported in the literature. T. S. Mahmood [25] proposed method for Cardiac Echo Videos for Decision Support systems. Cheolkon Jung and Joongkyu Kim [26] and Sunitha Abburu [27] reported text detection for golf and cricket matches respectively. Li Meng [28] proposed an algorithm for TV Commercial Detection.

Region based methods make use of low-level features, so these methods demonstrate elevated speed and achieve good results under simple background, but are sensitive to noises. Texture based methods give better results in complex backgrounds but computationally very expensive texture classification, which results in larger processing time. Due to very huge increase in multimedia data, Content based retrieval systems should be very efficient. That's why proposed system use region based segmentation method. Urdu is the type of Arabic type script. It is having different writing style and script. This paper focuses on the extraction of artificial Urdu text from video frames. Proposed method is tested on the Urdu news captured from different Urdu news channels.

III. PROPOSED METHODOLOGY

This paper exploits the elementary image processing tools along with the potential features for Urdu text extraction. Detailed architecture of proposed method is presented in Figure 2.

A. Pre-processing phase

Edge detection, binarization and dilation is done at pre processing stage. Edge detection has the fundamental importance in image analysis. Edge detection is used for the characterization of edges in an image and is a basic operation in segmentation, registration, and object identification. Results of object identification highly rely on accuracy of edge detection. Edge detection is the most appealing segmentation for the textual image, as edges are the vital characteristic of text objects. Different edge detection techniques are suitable for different applications. For specific applications edge detector may be tailored to take advantage of the domain knowledge. Proposed method use the vertical edges detected by Sobel operator for segmentation.

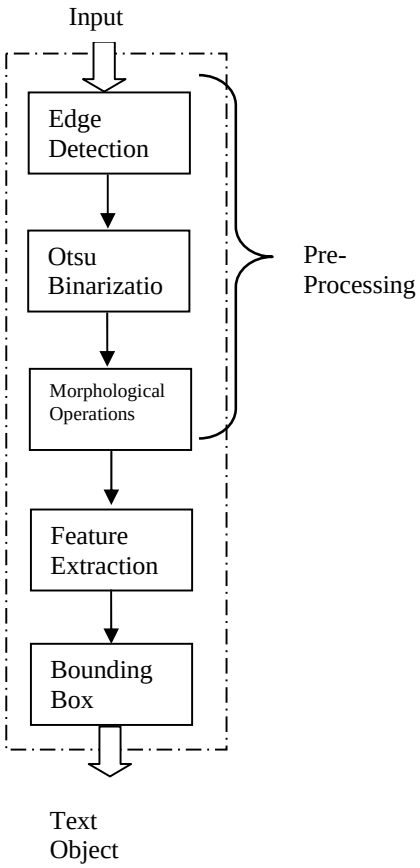

Figure 2: Flow chart of proposed technique

For efficient computation proposed algorithm work on the binary image. Image having edge map is converted to binary image using Otsu Binarization method [29]. After binarization, morphological treatment is carried out. Morphology is a extensive set of shape based image processing tools. A variety of tasks can be performed by variation of shape and size of morphological structures. Here, it is used for merger of neighboring characters to make the words and separate the words from the back ground. First dilation and then erosion is performed with different structuring elements. Image resolution varies a lot for images/videos. It can range from low resolution hand held device's camera to the professional ones. In order to deal with this resolution variation, structuring element for dilation should be adaptive. Size of the structuring element is calculated

$$L = 0.25 * S(1)$$

where S(1) is the width of the image.

L x 1 Structuring element is used for dilation to merge the neighboring characters and words.

B. Feature Extraction process

Text extraction is an area which is the focal point of researchers for many years because of its appliance in many

fields. Text extraction systems have many hereditary challenges that researchers have to face. Features, that can differentiate the text and non text objects, play a vital role in increasing the accuracy of text extraction process.

Many features has introduced in the literature. Proposed method defines set of features, suitable for Urdu text detection. Proposed feature set is composed of edge density, aspect ratio, horizontal fluctuation count (HFC) and vertical fluctuation count (VFC). These features are explained in the paper for completion.

Edge Density:

Edge density is the most widely used feature for text extraction. Normally text regions have more edges than the non text one. First any of the edge detection technique is applied then edge density is calculated by

$$\text{Edge Density} = \sum_{i=1}^{m}\sum_{j=1}^{n} E_{i,j}$$

Where,

$$E_{i,j} = \begin{cases} 1 & \text{if } im(i,j) = 1 \\ 0 & \text{otherwise} \end{cases}$$

and, $E_{i,j}$ is the edged image.

Aspect Ratio:

Many text extraction methodologies exploited this feature to reduce the false positives [6][7]. Aspect ratio can be defined as:

$$AR(c) = \min\{\frac{w(c)}{h(c)}, \frac{h(c)}{w(c)}$$

HFC and VFC:

Fluctuation count is the number of transitions from zero to one in a region. HFC and VFC are proposed in the [30]. Here these features are defined for completeness. Mathematical expression for Horizontal fluctuation count (HFC) is:

$$HFC = \sum_{i=1}^{m}\sum_{j=1}^{n} C_{i,j}$$

Where,

$$C_{i,j} = \begin{cases} 1 & \text{if } im(i,j-1) == 0 \text{ and } im(i,j) == 1 \\ 0 & \text{otherwise} \end{cases}$$

Vertical fluctuation count (VFC) is computed as:

$$VFC = \sum_{j=1}^{n}\sum_{i=1}^{m} D_{i,j}$$

Where,

$$D_{i,j} = \begin{cases} 1 & \text{if } im(i-1,j) == 0 \ and \ im(i,j) == 1 \\ 0 & \text{otherwise} \end{cases}$$

On the basis of above mentioned features, text blobs are identified and bounding box is generated. Step by step demonstration of proposed method is presented in figure 3.

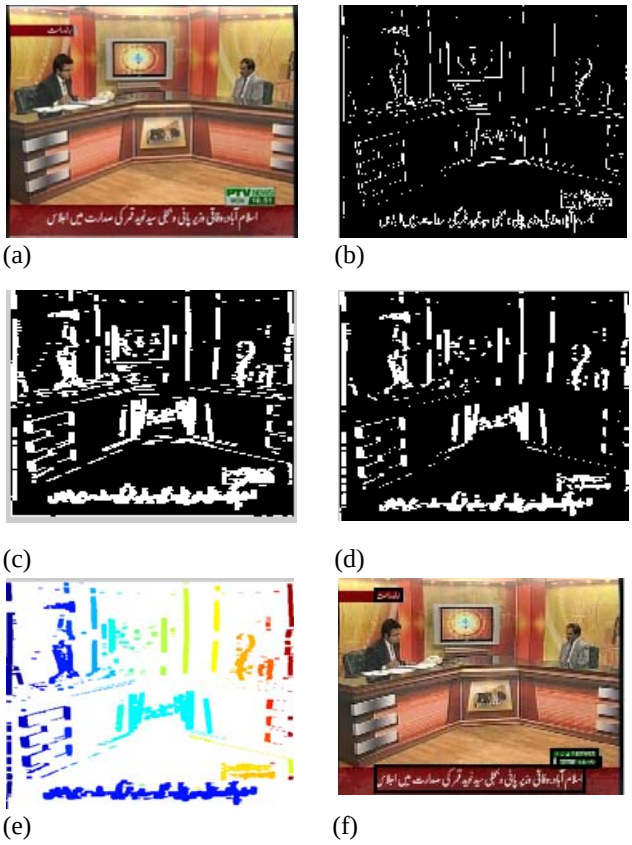

Figure 3: Demonstration of proposed method (a) Original Image (b) Edge map (c) Dilated Image (d) Eroded Image (e) Image with segmented objects (f) Text extraction result

IV. Experiments and Results

In order to evaluate the strength of above mentioned methodology, it is tested on Urdu news Dataset [31]. Dataset can be accesses through [32]. Dataset contains 1000 images and is divided into five categories i.e. News, sports, entertainment, business and religion. Dataset contains images with varying resolution and varying font sized text.

Results show the strength of the proposed method as Urdu text extraction tool. Precision Recall metric is used to measure the results of the proposed system. For the validation of the proposed work, area based definition of precision and recall is used. If G represents the ground truth text area and D represents the detected text area, then

$$\text{Recall} = \frac{G \cap D}{G}$$

$$\text{Precision} = \frac{G \cap D}{D}$$

Harmonic Mean is calculated by using following mathematical expression

$$H.M = 2\,\frac{P * R}{P + R}$$

Figure 4: Results of the proposed method

Table 1 summarizes the results of the proposed system and Figure 4 shows some of the results of the proposed method.

TABLE 1
RESULT OF THE PROPOSED METHOD

Precision	Recall	Harmonic Mean
0.79	0.74	0.76

Different categories of the dataset are tested separately and results are summarized in figure 5.

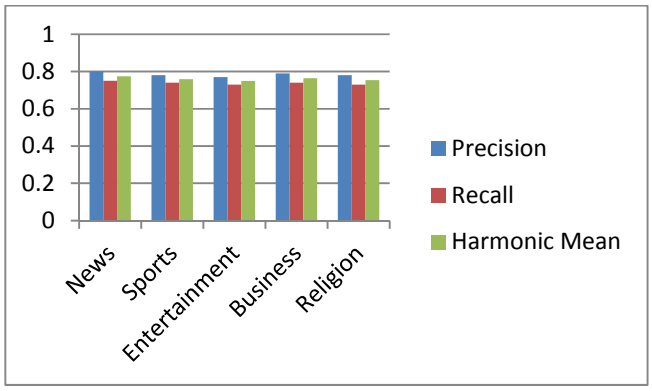

Figure 5: Performance of proposed method in different categories of dataset

Structuring element plays a vital role in text detection processes. Different sizes of structuring elements are investigated and compared with the adaptive structuring size. Results shown in figure 6 present difference between the adaptive structuring element and fixed ones.

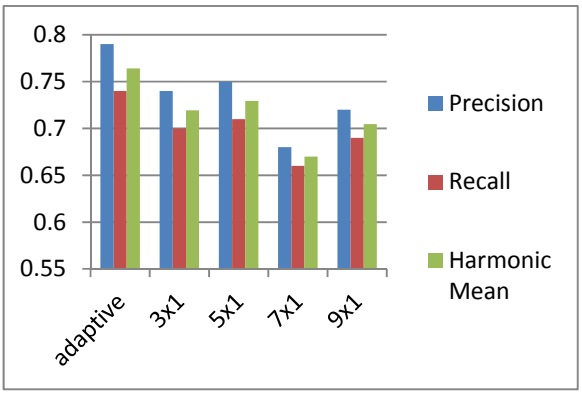

Figure 6: Performance of proposed method with different sizes of structuring element

V. CONCLUSION

This paper presents efficient Urdu text extraction method. Proposed method shows promising results and can deal with the variation in color, font and size of the text. It can also detect text in the images with variant resolution. This method can be extended to other languages having Nastaliq calligraphic style, such as Arabic.

REFERENCES

[1] H. P. Li, D. Doermann, and O. Kia. Automatic text detection and tracking in digital video. IEEE Trans. IP, 2000.

[2] K. I. Kim, K. Jung, and J. H. Kim. Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. IEEE Trans. PAMI, 2003.

[3] M. Zhao, S. T. Li, and J. Kwok. Text detection in images using sparse representation with discriminative dictionaries. IVC, 2010.

[4] K. Wang and S. Belongie. Word spotting in the wild. In Proc. ECCV, 2010.

[5] L. Neumann and J. Matas. A method for text localization and recognition in real-world images. In Proc. of ACCV, 2010.

[6] P. Shivakumara, T. Q. Phan, and C. L. Tan. A laplacian approach to multioriented text detection in video. IEEE Trans. PAMI, 2011.

[7] K. Jung, K. Kim, and A. Jain. Text information extraction in images and video: a survey. PR, 2004.

[8] J. Liang, D. Doermann, and H. Li. Camera-based analysis of text and documents: a survey. IJDAR, 2005.

[9] C.P. Sumathi, , T. Santhanam, and G.Gayathri Devi, “A SURVEY ON VARIOUS APPROACHES OF TEXT EXTRACTION IN IMAGES”, International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.4, August 2012.

[10] Lienhart, R., 2003. OCR, Video: A Survey and Practitioner’s Guide. Intel Corporation, Microprocessor Research Labs, Santa Clara, California, 155-184.

[11] Chuang, L., Ding, X., Wu, Y, 2006. An algorithm for text location in images based on histogram features and AdaBoost. J. Image Graph. 11(3), 325–331.

[12] Kim, K.I., Jung, K., Kim, J.H, 2003. Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. Pattern Anal. Mach. Learn. 26, 1631–1639.

[13] R. Lienhart and A. Wernicke, 2002. Localizing and Segmenting Text in Images and Videos, IEEE Transactions on Circuits and Systems for Video Technology 12(4), 256-268.

[14] J. Gllavata, E. Qeli and B. Freisleben, 2006. Detecting Text in Videos Using Fuzzy Clustering Ensembles, Proceedings of the Eighth IEEE Interanational Symposium on Multimedia, 283-290.

[15] D. Chen, J.M. Odobez and H. Bourlard, 2004. Text detection and recognition in images and video frames, Pattern Recognition, 595-608.

[16] J. Fabrizio, M. Cord, And B. Marcotegui(2009), “Text Extraction From Street Level Images,”, CMRT, Vol. Xxxviii, Part 3/W4 , pp. 199–204.
[17] Miriam Leon, Veronica Vilaplana, Antoni Gasull, Ferran Marques(2010),"Region-Based Caption Text Extraction",11th International Workshop On Image Analysis For Multimedia Interactive Services (Wiamis).
[18] B. Epshtein, E. Ofek, and Y.Wexler. Detecting text in natural sceneswith stroke width transform. In Proc. CVPR, 2010.
[19] L. Neumann and J. Matas. A method for text localization and recognition in real-world images. In Proc. of ACCV, 2010.
[20] Shuicai Shi, Tao Cheng, Shibin Xiao, Xueqiang Lv, 2009. A Smart Approach for Text Detection, Localization and Extraction in Video Frames. International Conference on Information Technology and Computer Science, 158 – 161.
[21] Jiangbo Xu, Xiuhua Jiang, Yuxia Wang, 2009. Caption Text Extraction Using DCT Feature in MPEG Compressed Video. World Congress on Computer Science and Information Engineering
[22] Shi Jianyong, Luo Xiling, Zhang Jun, “An Edge-based Approach for Video Text Extraction”, 2009 International Conference on Computer Technology and Development, 431 – 434.
[23] Xin Zhang, Fuchun Sun, Lei Gu, 2010. A Combined Algorithm for Video Text Extraction. Seventh International Conference on Fuzzy Systems and Knowledge Discovery, 2294 – 2298.
[24] Wang Zhiming, Xiao Yu, 2010. An approach for video-text extraction based on text traversing line and stroke connectivity”. International Conference on Biomedical Engineering and Computer Science, 1 – 3.
[25] Tanveer Syeda-Mahmood, David Beymer, Arnon Amir, 2009. Disease-Specific Extraction of Text from Cardiac Echo Videos for Decision Support. 10th International Conference on Document Analysis and Recognition, 1290 – 1294.
[26] Cheolkon Jung and Joongkyu Kim, 2009. Player Information Extraction for Semantic Annotation in Golf Videos. IEEE TRANSACTIONS ON BROADCASTING 55(1), 79 – 83.
[27] Dr. Sunitha Abburu, 2010. Multi level semantic extraction for cricket video by text processing. International Journal of Engineering Science and Technology 2(10), 5377-5384.
[28] Li Meng, Yong Cai, Min Wang, and Yuanxing Li, 2009. TV Commercial Detection Based on Shot Change and Text Extraction. 2nd International Congress on Image and Signal Processing, 1 – 5.
[29] N. Otsu, “ A threshold selection method from gray-level histograms”, IEEE Transaction on Systems, Man and Cybernatics, Vol.9, no 1 pp-62-66,1979.
[30] S. Tehsin, A. Masood, S. Kausar and Y. Javed, “Text Localization and Detection Method for Born-Digital images”, accepted in IETE Journal of Research, vol. 59, Issue 4.
[31] Imran Siddiqi, Ahsen Raza, “A Database of Artificial Urdu Text in Video Images with Semi-Automatic Text Line Labeling Scheme”, MMEDIA 2012 : The Fourth International Conferences on Advances in Multimedia
[32] https://sites.google.com/site/artificialtextdataset/urdu-script-1