(IJEACS) International Journal of Engineering and Applied Computer Science

Anatomical Survey Based Feature Vector for Text Pattern Detection

Samabia Tehsin
Department of Computer Science
Bahria University Islamabad Campus
Islamabad, Pakistan

Sumaira Kausar
Department of Computer Science
Bahria University Islamabad Campus
Islamabad, Pakistan

***Abstract*—The vital objective of artificial intelligence is to discover and understand the human competences, one of which is the capability to distinguish several text objects within one or more images exhibited on any canvas including prints, videos or electronic displays. Multimedia data has increased rapidly in past years. Textual information present in multimedia contains important information about the image/video content. However it needs to technologically verify the commonly used human intelligence of detecting and differentiating the text within an image, for computers. Hence in this paper feature set based on anatomical study of human text detection system is proposed.**

Keywords- Biologically-inspired vision; Content-based retrieval; Document analysis; Text extraction.

I. INTRODUCTION

There are very strong reasons to believe with substantial evidences that the first and foremost creation of universe would be a pen, which draws and writes. This writing develops into a plan to be manifested in shape of a logical working of every tier of all known and unknown galaxies. Even today in times of technological advancements and scientific progression, we could avoid only little about pen but not at all about writing. We write to read, teach, prove, plan, assimilate and disseminate. This all is worth proofing the very importance of alphabets, words, sentences and nevertheless, languages to understand.

In recent years there is a rapid increase in multimedia libraries, which raise the need of efficiently retrieving, indexing and browsing multimedia information. Several approaches have been introduced in the literature to retrieve image and video data. These techniques are based on color, texture, shape and relation between objects etc. For text based queries, text embedded in images and videos for retrieval is a very good option.

Visual texts appearing in multimedia data often impart knowledge about news headings, title of movie, brands of products, scores of a match, date and time when an event took place. All this information is vital for understanding and retrieving images and videos.

Text extraction and recognition process comprises of five steps namely text detection, text localization, text tracking, segmentation or binarization, and character recognition.

Text detection and localization are the primary steps in text extraction process. Different features are used in the literature to detect the text in the image. Mostly, these features are imported from other applications of computer vision and pattern recognition and these features are targeting the specific set of images. Most appropriate way of defining the text extraction features is to study the human brain operations and features used by humans to extract the text. No formal survey or study is conducted in the literature to study the human text detection system. This paper present the study of human text detection method and intelligent framework is presented to study the anatomical feature set used for text detection. Feature set describing the text objects in the image is also concluded.

The rest of the paper is organized as follows. Section II highlights some related work of the field. Section III proposes an intelligent framework to extract text detection features. Section IV presents the dataset used and results of the survey and section V provides some concluding remarks.

II. LITERATURE REVIEW

A variety of approaches of text extraction have been proposed during the past years. [1]-[10]. Comprehensive surveys can be found in [11]-[14]. But very less work has been done in defining the novel feature vectors for text detection. Most of the existing systems use few conventional features to classify text and non text objects. These features are generally defining few geometrical features of the text objects.

Zhong et al. [15] used a CC-based method using color reduction. They quantize the color space using the peaks in a color histogram in the RGB color space. Each text component goes through a filtering stage using heuristics, such as area, diameter, and spatial alignment.

Two geometrical constraints are applied by Wolf and Jolion [16] to eliminate the non text and detect the text objects from videos. One is the width to height ratio and the second one is number of text pixels of the component to area of the bounding box.

ISBN: 978-0-9957075-4-2 *DOI: 10.24032/ijeacs/0203/04*

Simple rules are used by Ezaki [17] to filter out the false detections. They imposed constraints on the aspect ratio and area to decrease the number of non-character candidates. Isolated characters are also eliminated from the text candidate list.

Hua et al. [18] used the constraints on height and width of the text candidates to reduce the false alarms. They also defined fill factor constraint to further reduce the non text objects. They defined the upper and lower limits for ratio of horizontal edge points to vertical edge points. They have also defined the upper limit for the ratio of edge points to total number of pixels in the area. Here the edge points represent horizontal edge, vertical edge and overall edge.

Epshtein et al. [19] present a novel image operator that seeks to find the value of stroke width for each image pixel, and demonstrate its use on the task of text detection in natural images. Many of the recent techniques are using this operator as part of text detection feature vector.

Local binary pattern is being used by Wei and Lin [20] for texture analysis. They first extracted the statistic feature of each text candidate by resizing each text candidate to 128x128 size. They then used Haar wavelet transform to decompose the text candidate to the four sub-band images including: low frequency (LL) band, vertical high frequency (LH) band, horizontal high frequency (HL) band and high frequency (HH) band. Next, they calculated the features in four sub-bands including mean, standard deviation and entropy of each sub-band. In addition to these statistic features, five features of the gray-level co-occurrence matrix (GLCM); energy, entropy, contrast, homogeneity and correlation, are calculated for each four direction in four wavelet sub-bands. 92-dimensional feature vector for each text candidate was generated, which was reduced to 36-dimensions using the principal component analysis (PCA).

After applying the morphological dilation on detected corner points in the image, [21] used five region properties as the features to describe text regions. These features are area, saturation, orientation, aspect ratio and position. The area is the foreground pixels in the bounding box. Saturation specifies the proportion of the foreground pixels in the bounding box that also belong to the region. Orientation is defined as the angle between x-axis and the major axis of the ellipse that has the same second-moments as the region. Aspect ratio of the bounding box is defined as the ratio of its width to its height. Position is defined by the region's centroid.

Shivakumara et al. [22] used two features to eliminate the false positives. One is the straightness and the other one is The first feature, straightness, comes from the observation that text strings appear on a straight line (their assumption), while false positives can have irregular shapes. The second feature, edge density, is defined as the ratio of edge length to the connected component area. Ranjiniand and Sundaresan [23] used the area to find the text area blob.

There is a need of in-depth study of text structures. Anatomical study of human text detection can be useful for identification of such features. And there is also a need to mathematically model those bio inspired features to make it workable for machines. Detailed geometrical and statistical study of text objects is also required.

III. Methodology

Humans are very good in detecting text in the images and scenes around them. So it's very important to study the anatomical text detection system before deciding the features for machines, to detect text. This is the reason that the intelligent framework to study the human text detection system is presented and testified in this paper.

People won't be able to answer if direct investigation about the features of text and nontext objects is carried out. So an indirect framework is designed so that features can be summarized. First the dataset containing the text images is collected with variation in font, color, size and language of the text.

TABLE 1. Parameters of proposed framework

Parameter	Possible Values	
Language	Known	Unknown
Regularity	Symmetric	Non symmetric
Density	Single	Group
View	Distant	Close

Framework is designed by observing the little child who is yet in age of learning any specific language of world. Till he does not know how to read and write any language, he does not have even the idea of difference between a drawing and writing, but then he gets to know spelling of words, way of writing a specific language and development of words into sentences. Actually this is the time he starts differentiating between what is drawn and what is written, though he yet does not know, how to read or write all known languages less one, he is taught. But keeping in mind the way of writing and specific texture of words, running in a regular scheme as a reference, he can make out that the one displayed in front of him is a text and not a simple drawing. So he can detect and differentiate now, though yet cannot assimilate to read.

Based on this observation, it is deduced that recognition and detection are two different processes. In order to get the separate observations for detection and recognition, language is considered as the parameter in the framework. So text of known as well as unknown language is added to the dataset. In order to emphasize this point, two views are introduced in the framework. One is the close view, that is distance between the viewer and the text image is very less, that the person is able to see and assimilate the text. Other one is the distant view, in this the person can see but not able to assimilate the text .e.g. small

ISBN: 978-0-9957075-4-2 DOI: 10.24032/ijeacs/0203/04

text appeared on billboard viewed from the distance can be seen but can't be read.

Other two parameters are the symmetry and density of words. Symmetric text is the one having same height, size, font and color for all characters, aligned in some specific direction. Non symmetric text misses all or some of the features of symmetric text.

(a)

(b)

Figure 1. (a) Isolated Characters (b) Text in monograms (Irregular text)

Four parameters and their possible values defined in the framework to explore the text features are shown in table 1.

For the sake of clarity, parameters are defined here. Known language is the language whose writing style is familiar to the viewer and unknown language is one with the alien writing style. If a person can read and write the English language only, the Spanish and Korean would be unknown language for that person. Regular text is the text with symmetry e.g. news credits and irregular text is the text without symmetry, which normally appears in the monograms and product labels. Examples of different text groups of images are presented in fig. 1. View is defined in terms of distance of viewer from the text. Close view is the one, text can be seen and read comfortably and it is distant view if text can be seen but cannot be read.

Twelve different groups are formulated by combination of above mentioned parameters. These groups can be observed through the parameter tree shown in fig. 2. Alphabetical tag (A-L) is attached to each of these groups. For example, Label 'A' is assigned to the group with isolated character of a known language, which is viewed from distance.

IV. DATASET AND RESULTS

A. *Dataset*

Dataset includes images with variety of text styles. Dataset includes text with variation in font, style, color and language.

Dataset also include the artistic text, text in monograms, isolated and grouped characters.

Dataset is divided into six test sets; these are:

A. Isolated characters of known language
B. Isolated characters of unknown language
C. Grouped characters of known language, with symmetry
D. Grouped characters of known language, without symmetry
E. Grouped characters of unknown language, with symmetry
F. Grouped characters of unknown language, without symmetry

Twenty images of each category are included in the dataset; which makes the total of 120 images in the dataset. Some images from the dataset are presented in the fig. 3

ISBN: 978-0-9957075-4-2 DOI: 10.24032/ijeacs/0203/04

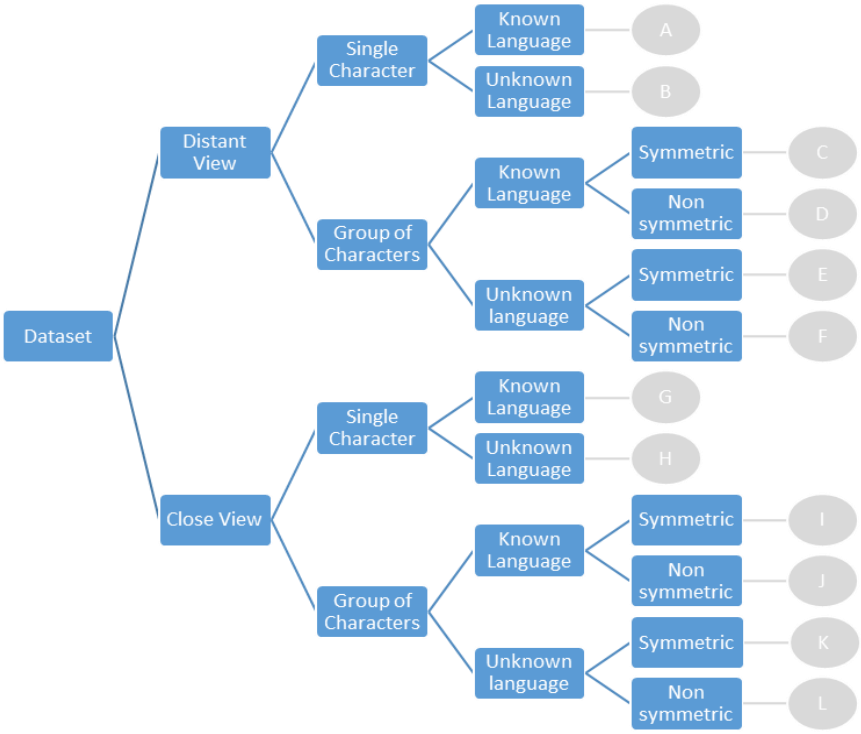

Figure 2. Parameter Tree

Figure 3. Sample dataset images

B. *Participants*

Fifty individuals, ranging in age from 7 to 12 years, completed the survey. This age group is chosen because;

ISBN: 978-0-9957075-4-2 DOI: 10.24032/ijeacs/0203/04

people of this age group usually have reading and writing knowledge of only one language. The people below this range may not have knowledge of reading and writing any language. People age above this age group may have knowledge of many languages or they have developed the knowledge of relating different facts to detect text in the image.

C. Results

Six test sets of dataset are checked under two conditions; distant view and close view. Parameter tree in fig. 2 describes all the possible test cases for the experimentation. Dataset categories (1-6) for distant view is represented by (A-F) in parameter tree and (G-L) characterize close view cases.

Table 2 shows the results of the carried out experiments. Test Case shows the labels associated to each case in the parameter tree. Next four columns represent different possible values for the parameters of framework. Last column represents the accuracy of the detection. Accuracy is computed by

$$\text{Accuracy} = \frac{\text{Correctly detected objects}}{\text{Total number of text objects}} \text{ X } 100$$

TABLE 2. Results of the experiments for different values of the parameter

Test Case	View	Density	Language	Symmetry	Accuracy
A	Distant	Single	Known	-	28%
B	Distant	Single	Unknown	-	17%
C	Distant	Group	Known	Symmetric	99%
D	Distant	Group	Known	Non-Symmetric	39%
E	Distant	Group	Unknown	Symmetric	97%
F	Distant	Group	Unknown	Non-Symmetric	35%
G	Close	Single	Known	-	100
H	Close	Single	Unknown	-	34
I	Close	Group	Known	Symmetric	100
J	Close	Group	Known	Non-Symmetric	99
K	Close	Group	Unknown	Symmetric	99
L	Close	Group	Unknown	Non-Symmetric	41

Figure 4 shows the result of the experimentation. In the figure horizontal legends show the different test sets defined in the dataset section. Vertical legends show the accuracy of the detection process. Experiments show that test sets three and five give very good detection results both for distant and close view. Common parameters between these test cases were the grouped characters and symmetry of text. It means if the characters are grouped and are in some symmetry, text will get detected whether seen closely or from distance. In other words, text would get detected whether recognized or not.

Text detection gives poor results for category four and six, except when of known language and viewed from nearby. It means if symmetry is not there in text, it won't be detected unless recognized.

From the study of test cases one and two, it is clear that isolated characters are detected only if of known language and viewed closely.

From the above discussion it can be concluded that text detection and recognition are two different stages for human text vision system. Text can be detected in following conditions:

- Text should be symmetric, that is it should have equal gaps between the characters and words and height of the characters should be approximately same.
- Text should be in groups i.e. it should be combination of three or more characters

ISBN: 978-0-9957075-4-2 DOI: 10.24032/ijeacs/0203/04

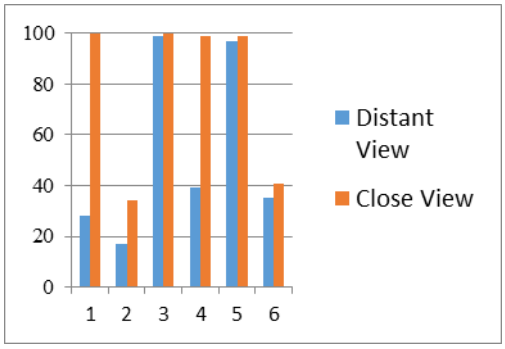

Figure 4. Results of experiments

If above two features exists in the text that can be detected from distant as well as from close view. It will be detected whether it is of known language or unknown language. But if text lacks the above mentioned features, it can only be detected if recognized. So the above mentioned features are necessary to detect the text without recognition i.e. without knowing the shape of the alphabets of language.

So in order to detect the text before the recognition process, symmetry of the text can be exploited. This symmetry may includes

- Periodic gaps between characters.
- Even distribution of foreground object.
- Constant instantaneous height of text object.
- Ratio of background and foreground remains almost same throughout the object.

These features can be used for the text detection process. So there is a need to develop the mechanism for regular and irregular texts separately. Humans can even detect the text of unknown language, where they cannot recognize the characters and words. However the criterion remains, that common features between languages, that is a regular scheme of words and texture prevails. Whereas to read a specialized text, like monograms and logos, where texture of words and sequence may vary, one may know the difference of a drawing or a text only when he knows the specific language in which it is written. This means, that to read an uncommon and irregular writing, the brain of human needs to assimilate first, and to read afterwards.

V. Conclusion

In this paper an intelligent framework is designed to explore the features, used by human to detect text in the image. This framework consists of four parameters, each having two possible values. Total twelve test cases can be formulated with the combination of these parameters. These test cases are tested by the fifty observers and it is deduced that text detection and recognition are two separate steps in the human text detection system. In some cases detection is carried out without recognition and in other cases detection is done through the recognition. In the later cases detection won't be possible without recognition.

From the experimentation it is observed that if detection has to be carried out without the recognition process; following features should exist in the text; one is symmetry and other one is the group of characters. If either of the features is missing from the text, text cannot be detected unless recognized.

Mathematical representation of feature vector can also be formulated as future research work.

References

[1] H. P. Li, D. Doermann, and O. Kia. Automatic text detection and tracking in digital video. IEEE Trans. IP, 2000.
[2] Tehsin, Samabia, et al. "Text Localization and Detection Method for Born-digital Images." IETE Journal of Research (Medknow Publications & Media Pvt. Ltd.) 59.4 (2013).
[3] 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.
[4] Tehsin, Samabia, et al. "Fuzzy-Based Segmentation for Variable Font-Sized Text Extraction from Images/Videos." Mathematical Problems in Engineering2014 (2014).
[5] M. Zhao, S. T. Li, and J. Kwok. Text detection in images using sparse representation with discriminative dictionaries. IVC, 2010.
[6] K. Wang and S. Belongie. Word spotting in the wild. In Proc. ECCV, 2010.
[7] Tehsin, Samabia, et al. "A CAPTION TEXT DETECTION METHOD FROM IMAGES/VIDEOS FOR EFFICIENT INDEXING AND RETRIEVAL OF MULTIMEDIA DATA." International Journal of Pattern Recognition and Artificial Intelligence (2014).
[8] L. Neumann and J. Matas. A method for text localization and recognition in real-world images. In Proc. of ACCV, 2010.
[9] P. Shivakumara, T. Q. Phan, and C. L. Tan. A laplacian approach to multioriented text detection in video. IEEE Trans. PAMI, 2011.
[10] Samabia Tehsin, and Asif Masood (2014), "Geometrical Analysis Based Text Localization Method", Int'l Conf. IP, Comp. Vision, and Pattern Recognition (IPCV'14), pp 554-560
[11] K. Jung, K. Kim, and A. Jain. Text information extraction in images and video: a survey. PR, 2004.
[12] J. Liang, D. Doermann, and H. Li. Camera-based analysis of text and documents: a survey. IJDAR, 2005.
[13] 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.
[14] Samabia Tehsin, Asif Masood and Sumaira Kausar, " Survey of Region-Based Text Extraction Techniques for Efficient Indexing of Image/Video Retrieval", International Journal of Image, Graphics and Signal Processing, 2014, 12, pp 53-64
[15] Zhong, Y., Karu, K., & Jain, A. K. (1995). Locating text in complex color images. Pattern recognition, 28(10), 1523-1535.
[16] Wolf, C., & Jolion, J. M. (2004). Extraction and recognition of artificial text in multimedia documents. Formal Pattern Analysis & Applications, 6(4), 309-326.
[17] Ezaki, N., Kiyota, K., Minh, B. T., Bulacu, M., & Schomaker, L. (2005, August). Improved text-detection methods for a camera-based text reading system for blind persons. In Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on (pp. 257-261). IEEE.
[18] Hua, X. S., Chen, X. R., Wenyin, L., & Zhang, H. J. (2001, September). Automatic location of text in video frames. In Proceedings of the 2001 ACM workshops on Multimedia: multimedia information retrieval (pp. 24-27). ACM.
[19] Epshtein, B., Ofek, E., & Wexler, Y. (2010, June). Detecting text in natural scenes with stroke width transform. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (pp. 2963-2970). IEEE.
[20] Wei, Y. C., & Lin, C. H. (2012). A robust video text detection approach using SVM. Expert Systems with Applications, 39(12), 10832-10840.

ISBN: 978-0-9957075-4-2 DOI: 10.24032/ijeacs/0203/04

[21] Zhao, X., Lin, K. H., Fu, Y., Hu, Y., Liu, Y., & Huang, T. S. (2011). Text from corners: a novel approach to detect text and caption in videos. Image Processing, IEEE Transactions on, 20(3), 790-799.
[22] Shivakumara, P., Phan, T. Q., & Tan, C. L. (2011). A laplacian approach to multi-oriented text detection in video. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 33(2), 412-419.
[23] Ranjini, S., & Sundaresan, M. (2013). Extraction and Recognition of Text From Digital English Comic Image Using Median Filter. International Journal.

AUTHORS' PROFILES

Samabia Tehsin did her Ph.D. and M.S. in Image Processing from MCS, NUST. Currently she is working in Bahria University Islamabad as an Asst. Professor. Her areas of research are Digital Image processing, computer Vision and Document Analysis.

Sumaira Kausar is an Asst. Professor at Bahria University Islamabad. Her research interests are Digital image Processing, Gesture recognition and machine learning.