International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 22 • 2017

Text Extraction from Image of Word Cloud Using Tries Data Structure

Lalita Kumari[a] J.L. Raheja[b] and Vidyut Dey[c]

[a]Department of Electronics & Communication Engineering, NIT Agartala, Tripura, India
E-mail: kumaril2003@yahoo.co.in
[b]Digital System group, CSIR/ CEERI, Pilani, Rajasthan, India
E-mail: jagdish.raheja.ceeri@gmail.com
[c]Department of Production Engineering, NIT Agartala, Tripura,India
E-mail: vidyut.pe@nita.ac.in

Abstract: This paper presents a new method for text extraction from word cloud images. In order to recognize characters, edge mapping was carried out with the help of a 2D-DWT which helped to perform multi-resolution analysis too. In order to reduce the number of training samples k-mean clustering algorithm and support vector machine (SVM) was used. Since, the characters in a word cloud image are scattered in various directions at any moment of time, even after character recognition, word formation remain a challenging task. In order to extract a lexical text from the extracted characters, tries data structure was used. The proposed method is able to detect/extract text characters and also can successfully recognize words. Experimental result confirms the usability and effectiveness of the proposed algorithm.
Keywords: *Text extraction, word cloud, tries data structure, SVM, k-mean clustering, DWT.*

1. INTRODUCTION

The task to detect text from a video is far more challenging than detecting texts from a static image. Besides, another challenge has emerged in the area of extracting meaningful text from a word cloud or from among a cluster of jumbled characters [4][12]. Word cloud can be tweaked with different color, layout and font schemes. It gives greater prominence to those words that appear more frequently in the scene text. Word clouds are used for brain storming. It helps people to distill summarized information. Not only can these, word clouds help to assimilate important ideas, and concepts quickly. It also helps to improve vocabulary. Word cloud may also be used as a word bank for any individual topic.

Fig. 1 represents few samples of word-cloud. These types of word-cloud are messy and generally look as connection graph. If one traverses along few specific directions the graph generates meaningful words and rest is meaningless. These connection graphs sometime give meaningful word but they don't give any information about precedence order of characters. Therefore dictionary help is essential to recognize character precedence order and graph traversal path identification.

In this paper we have presented an algorithm to extract text from word cloud, jumbled word from scene image as well as born-digital image. The proposed method uses tries data structure for word extraction.

(a) (b) (c)

Figure 1: Sample word cloud image

2. RELATED WORK

One has to go through several steps, in a proper sequence, in order to extract text from a natural scene image or a video image [1][2][3][7]. Text characters are detected through Maximally Stable Extremal Regions (MSER) [5][14], connected component cluster analysis. Sometimes after detection of a probable area containing text, within an image, these areas are further refined with some machine learning approaches to distinguish between text/non-text [2][8]. This helps to extract text characters. These text characters, if are arranged in a sequential manner, along a predefined direction forms a word.

Roy[2] used Bayesian classifier in combination with wavelet decomposition for recognizing text in video frames. This method was able to detect multi-oriented text in image and video sequence frames. Khare in [3] presented a histogram oriented moment descriptor for text detection in video frames. On the basis of histogram analysis of connected components, text or non-text was detected. Anthimopoulos [9] used random forest classifier and a multi-level adaptivecolor edge local binary pattern for text detection from images and video. Lots of research work has been carried out for text extraction on different kind of images and video frames. Some of them are very much effective to extract text when in a simple line but when the characters are jumbled in multi-direction they fail. Therefore specialized method is required to recognize word correctly from such situation. The authors in this paper present a novel two phased approach by using 2-D DWT, k-mean Clustering, SVM, and Tries to detect and extract word correctly from image of word-cloud and crosswords or puzzle words where characters are jumbled.

3. PROPOSED METHOD

The proposed method is divided into two phases, first one recognizes characters from an image, and second part extracts words correctly from the word-cloud. First phase uses two dimensional discrete wavelet transform (2-D DWT) for edge mapping and multi-resolution analysis followed by k-mean clustering [13]. The result is further refined by using support vector machine (SVM)[15]. Second phase starts with character recognition using OCR pattern analysis and location mapping/binding with recognized characters within the image file. This location information is stored as vector and tries data [4][16] structure is used to recognize words correctly irrespective of word orientation in the word cloud. Generalized flowchart of the proposed algorithm is presented in Fig. 2

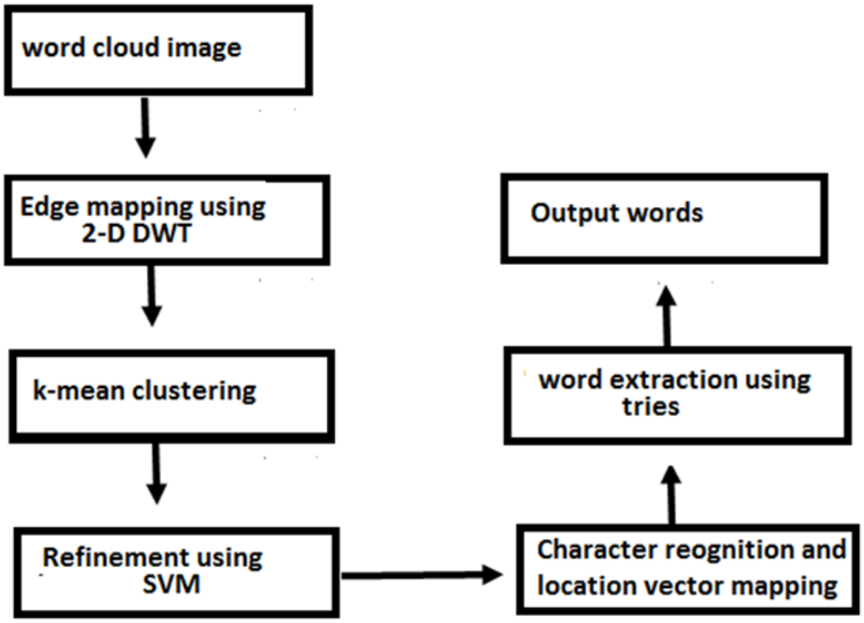


Figure 2: Flowchart of proposed algorithm using use tries

3.1. Discrete Wavelet Transform

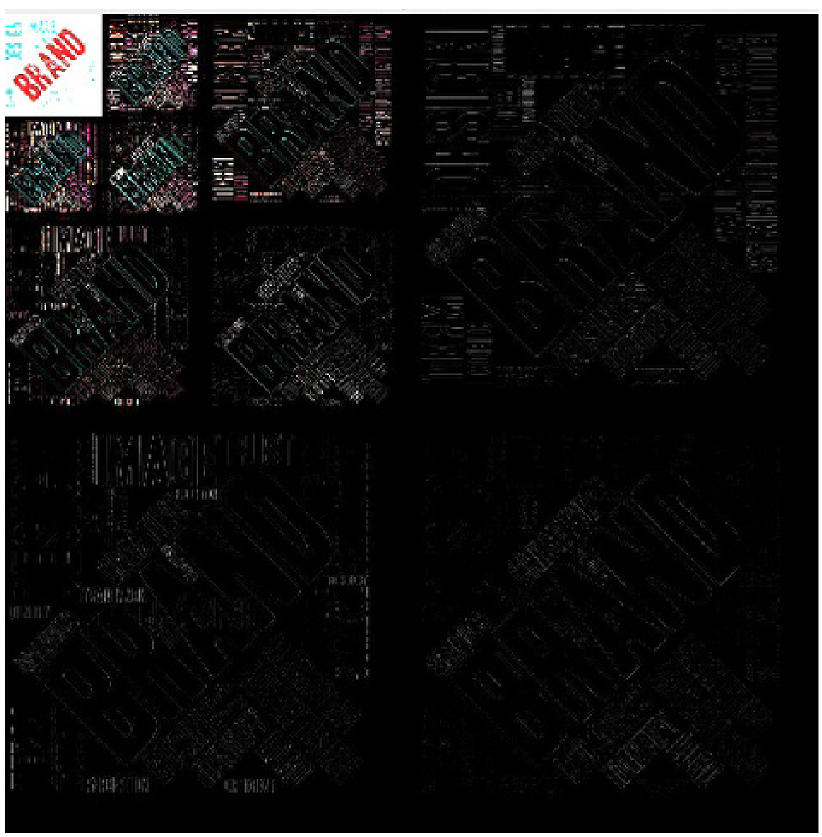


Figure 3: 3-level of discretewavelet transform

As wavelet transform provide localization information by breaking the signal and passing it through a series of filters. Its discrete version *i.e*. Discrete Wavelet Transform (DWT) is very much useful to analyze image data both in terms of frequency and location information. Details of the coefficients and approximation coefficients of signal X is represented by equation-1 and equation-2 respectively. The symbols '*g*' and '*h*' used in the equations represent low pass filter and high pass filter respectively. Fig. 3 displays output of 2D-DWT operation on input image 1.a) up to 3 level.

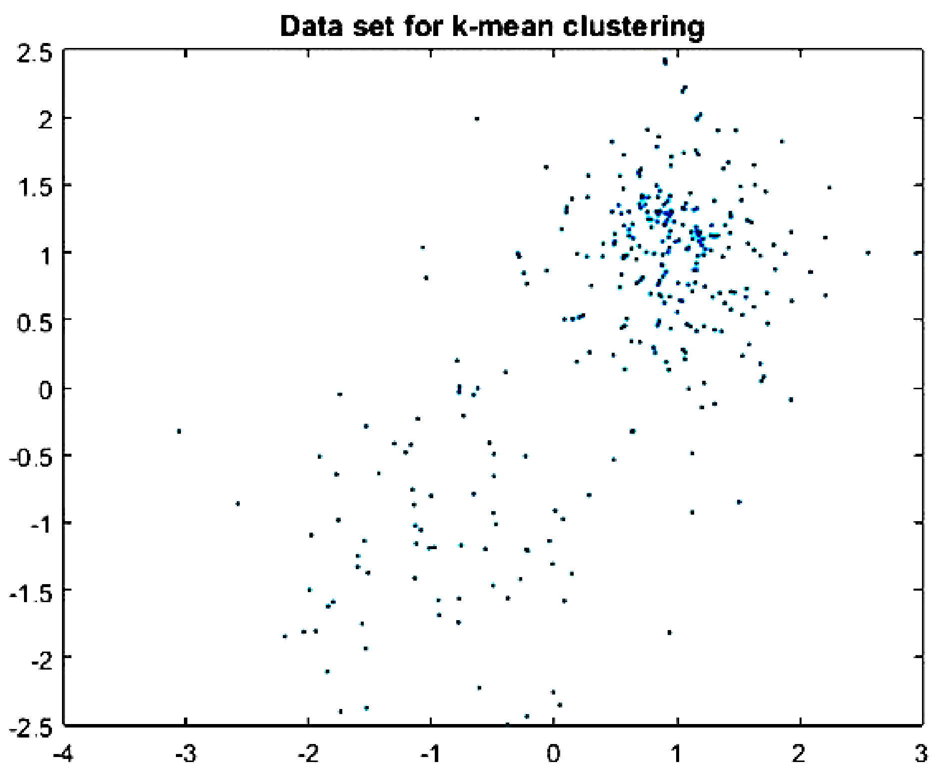
$$Y_{low}[n] = \sum_{-\infty}^{\infty} X[k]g[2n-k] \quad (1)$$

$$Y_{high}[n] = \sum_{-\infty}^{\infty} X[k]h[2n-k] \quad (2)$$

3.2. K-means clustering

After 2-D DWT, *k*-mean clustering is used to group dataset into non-overlapping set. These non-overlapping sets are identified using similarity of properties such as Euclidean distance. Basic steps of K-means clustering algorithms are:

1. Select n random initial centroid for clusters
2. Calculate distance of each data points and reassign to their closest centroid
3. Compute centroid for newly constructed clusters
4. Repeat step 2 and 3 until centroids do not change.



(a)

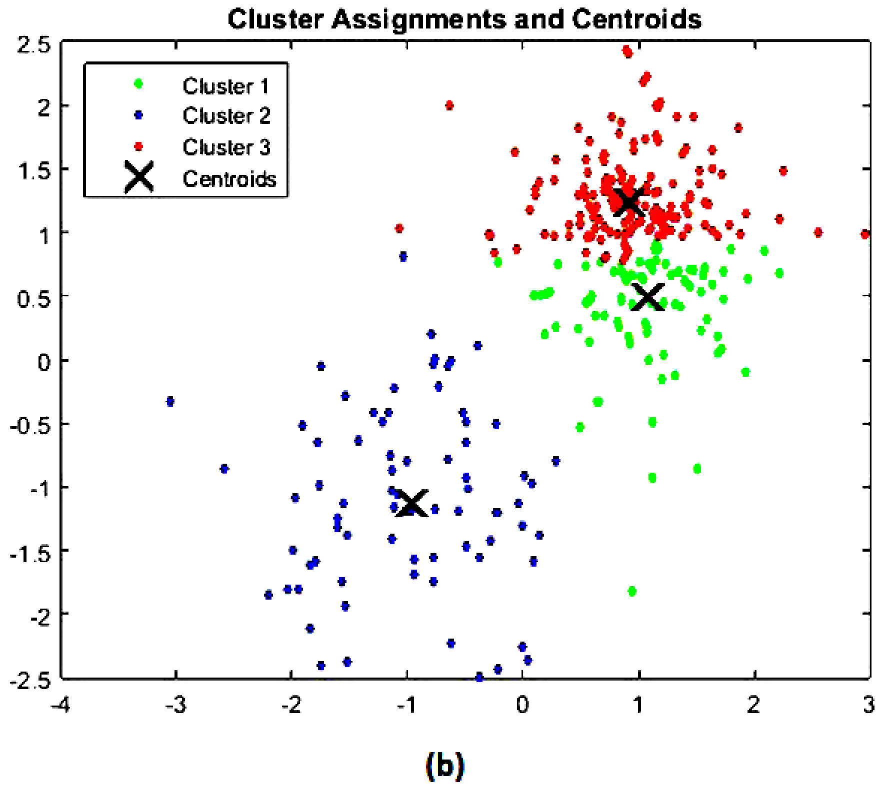


Figure 4: Cluster formation and centroid computation

Figure 5: Representation of cluster index and different clusters

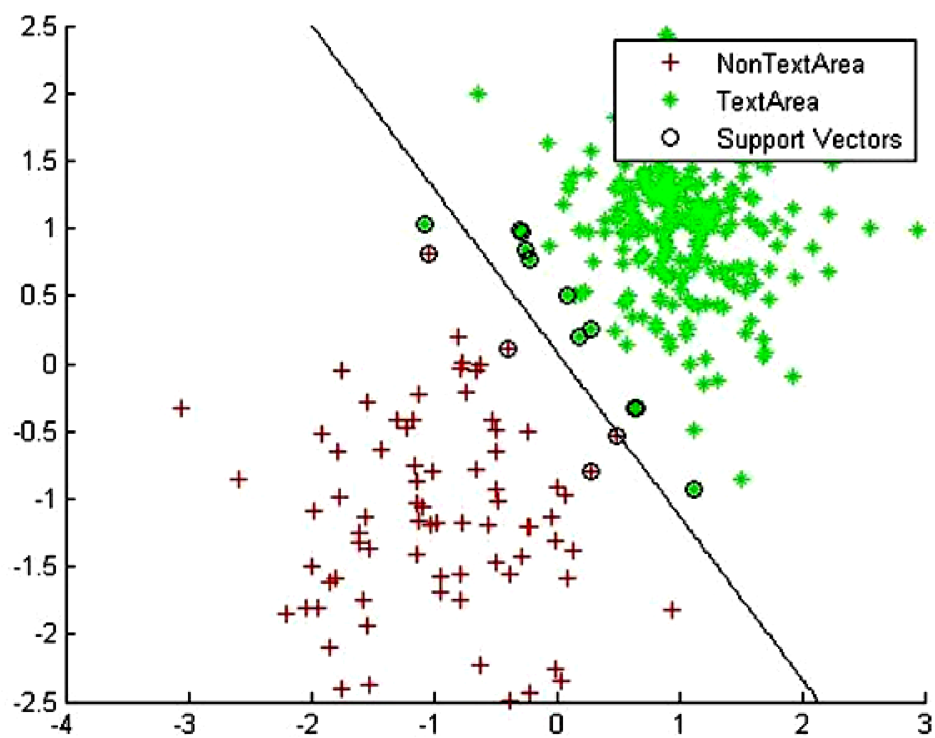
Objective of *k*-means algorithm is to partition dataset into groups C1,C2,….,Cn such that centroid of C_i, (represented by μ in equation 3) be minimum. Objective of *k*-means algorithm is to find μ_i, such that G (represented by equation 4) beminimum. K-mean clusters for three chosen centroids are displayed in Fig.4. Fig. 4(*a*) represents data points and Fig. 4(*b*) represents finally obtained three centroids after number of iterations. Fig. 5 displays cluster index and different clusters obtained from input image 1(*a*) displaying text objects inside it.

$$\mu(C_i) = \underset{\mu}{\arg\min} \sum_{\forall X \in C_i} \text{Euclidean Distance}(x, \mu) \tag{3}$$

$$G = \min_{\mu 1, \ldots \mu k} \sum_{i=1}^{k} \sum_{\forall X \in C_i} \text{Euclidean Distance}(x, \mu i) \tag{4}$$

3.3. SVM Classifier

As recognized text area from input image is categorized into exactly two groups i.e. either it will belong to text area or not, it is further refined using support vector machine. SVM classifier is used for feature extraction, classification, and text characterization. For this, a training data has been used to create classifier structure. Fig.6a represent SVM classifier on training data set and Fig.6b display prediction of text and non-text area based on training data. Using SVM classifier extracted text is further refined for greater accuracy. After this approach characters are recognized separately using optical character recognition [10] techniques.



(a)

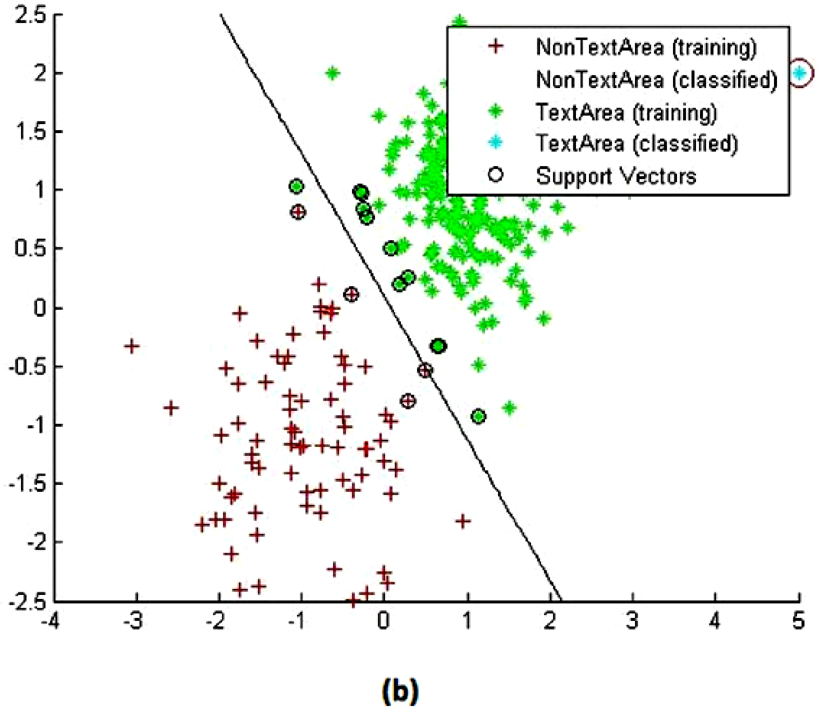


Figure 6: SVM classifier on training data set prediction

3.4. Word recognition from mess of characters using tries

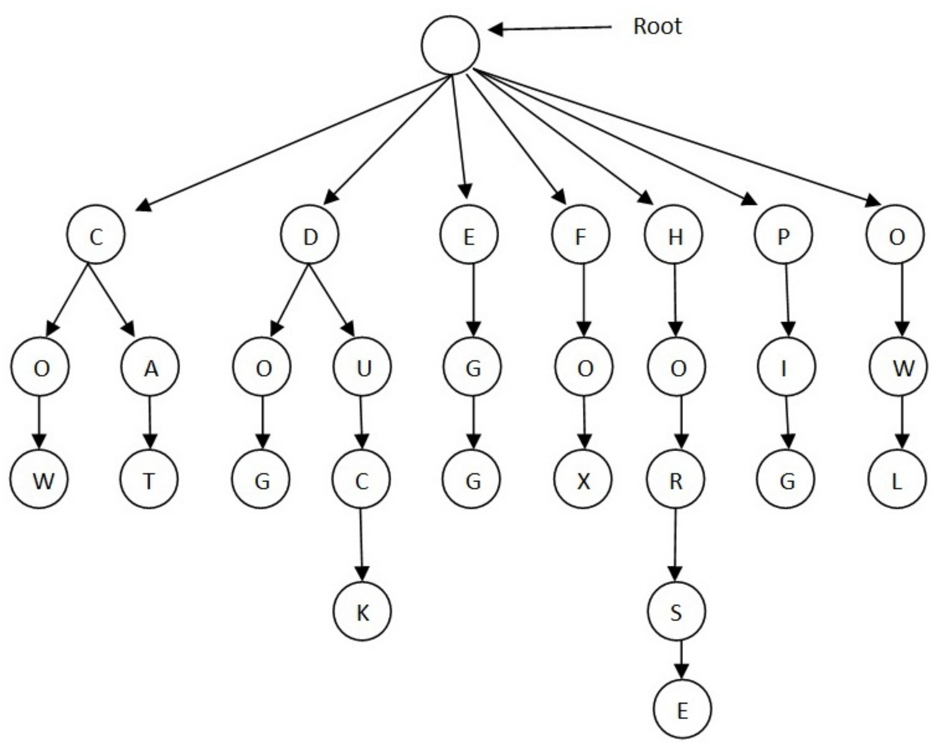


Figure 7: Tries structure generated for Fig. 1*b* on closed domain

As we have set our input image as word-cloud, word recognition becomes challenging task even after character recognition step. It is not easily possible to group a set of characters to form a word because orientation of words varies largely on same documents. Sometimes single character requires to be grouped into multiple words set to form word correctly. To solve this problem we have implemented tries data structures which work on limited input word from closed domain. Working of tries on limited words from closed domain is shown in Fig.7. Steps of Word recognitions from meshed characters are as follow:

1. Calculate total count (*n*) of appearing characters.
2. Construct vector of size n and store centroid of each characters, as location of characters.
3. Compute neighboring centroid for each saved centroid of character.
4. Generate tries for limited words of specific domain.
5. For each character *c*, select its entire neighbor T.
6. For all $t \in$ T, search through tries and change to next node.
7. Set *t* into *c* and move to step 3.

4. RESULT ANALYSIS AND CONCLUSION

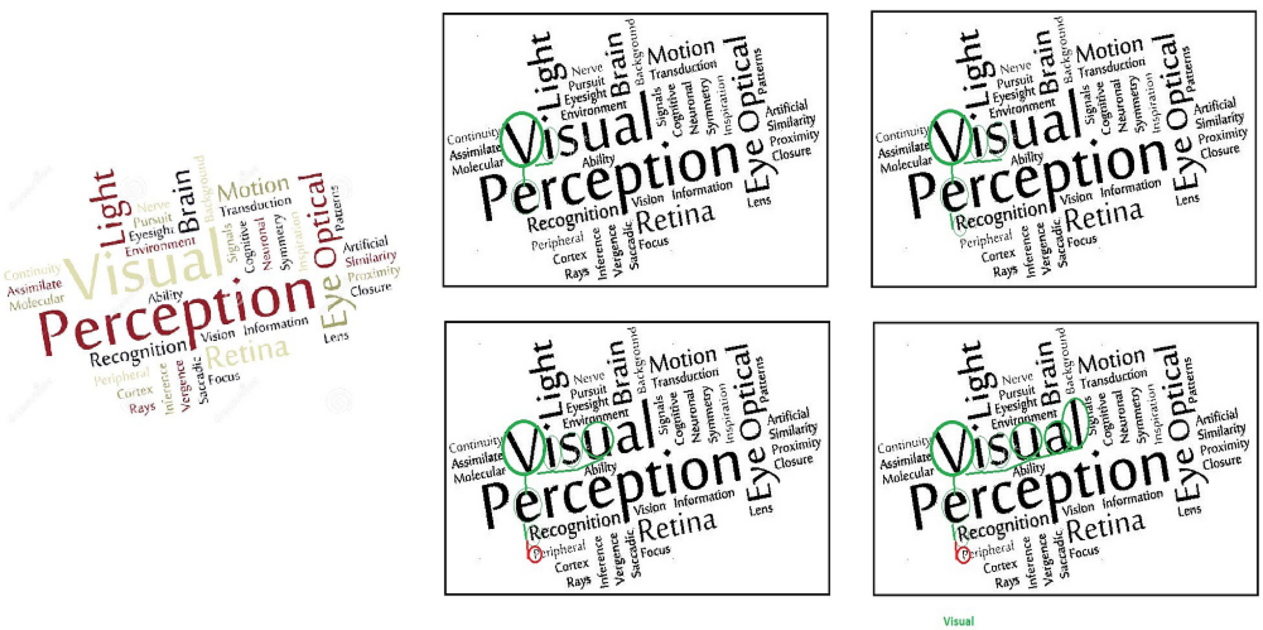


Figure 8: Word generation using tries (*a*)

In order to carry out the experiments for result analysis, we have used born digital test image, word cloud images, crosswords and puzzle images in both categories *i.e* created digital image, and captured image via camera. Fig.8 and Fig.9 display resulting text extracted after tries data structure implementation above the extracted characters vectors. The authors thus presented an approach to detect and extract word correctly from image of word cloud and crosswords or puzzle words where characters are jumbled. The authors tested the proposed technique on several images including born-digital image as well as camera captured images. The images were sub-categorized images into three subgroup (1) normal text image, (2) cross-word type images, and (3) word cloud images. Table1 shows that normal text type images give same result on with and without use of tries but in case of crossword and word-clouds, result accuracy increase significantly on using tries.

The proposed method, thus,is very much effective on word extraction correctly from word-cloud.

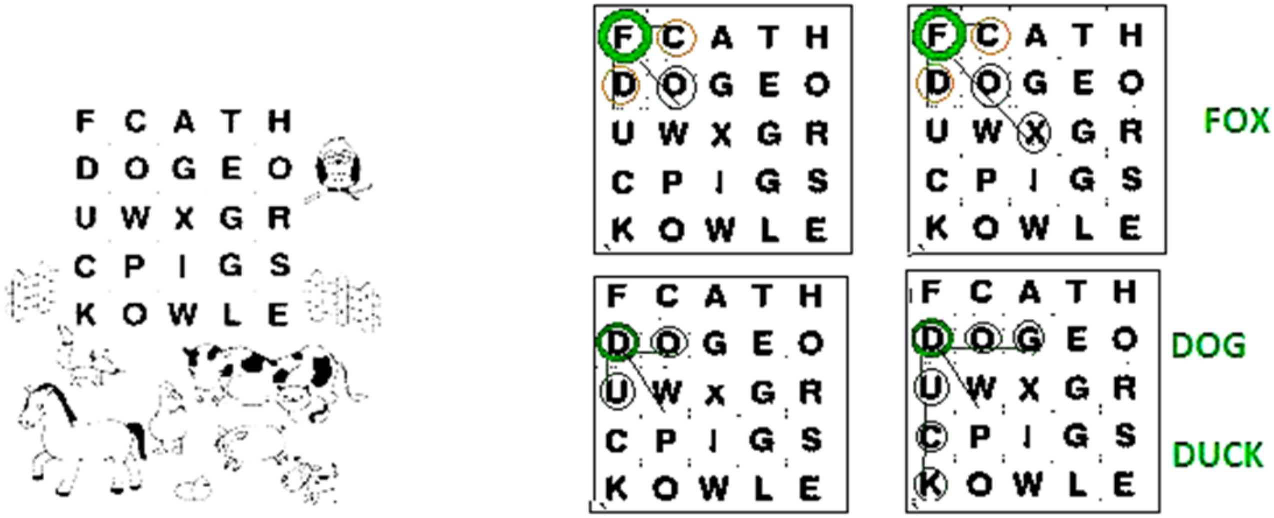


Figure 9: Word generation using tries (*b*)

Table 1
Result Comparison

Image type		*Word Recognition accuracy*	
		Without Tries	*With Tries*
Born Digital image	Aligned text	0.90	0.90
	Cross word	0.05	0.72
	Word cloud	0.40	0.65
Camera Captured Image	Aligned text	0.86	0.86
	Cross word	0.05	0.70
	Word cloud	0.38	0.61

REFERENCES

[1] L. Kumari, J.L. Raheja , V. Dey , “Text Detection from Live Sports Video, in International Journal of Control Theory and Applications vol 9, issue 20, pp-411-419 (2016)

[2] S. Roy, P. Shivakumara, P.P Roy, U. Pal, C. L. Tan, T. Lu “Bayesian classifier for multi-oriented video text recognition system, in Expert Systems with Applications, vol 42 pp.5554-5566 (2015)

[3] V. Khare, P. Shivakumara, P. Raveendran “A new Histogram Oriented Moments descriptor for multi-oriented moving text detection in video, in Expert Systems with Applications vol 42 pp.7627-7640 (2015)

[4] M. Frunzeanu “Using wikis, word clouds and web collaboration in Romanian primary schools, in The 6th International Conference Edu World 2014 Education Facing Contemporary World Issues, 7th - 9th November 2014

[5] Gomez, L., Karatzas, D. “MSER-based real-time text detection and tracking”, in Proceedings of ICPR pp. 3110-3115 (2014).

[6] Robert J. Riggs, S. Jack Hu “Disassembly liaison graphs inspired by word clouds, in Forty Sixth CIRP Conference on Manufacturing Systems 2013

[7] Shivakumara, P., Phan, T. Q., Shijian, L., Tan, C. L., "Gradient vector flow and grouping based for arbitrarily-oriented scene text detection in video images" in IEEE Transactions on Circuits and Systems for Video Technology, pp 1729-1739 (2013).

[8] Roy, S., Roy, P., Shivakumara, P., Pal, U. " Word recognition in natural scene and video images using hidden markov model". In Proceedings of the NCVPRIPG pp.14 (2013).

[9] Anthimopoulos, M., Gatos, B. "Detection of artificial and scene text", in images and video frames. Pattern Analysis and Applications, pp 431-446.(2013).

[10] Fernandez-Caballero, A., Lopez, M.-T., Castillo, J.-C. "Display text segmentation after learning best-fitted OCR binarizationparameters", in Expert Systems with Applications, pp4032-4043 (2012).

[11] Shahab, A., Shafait, F., Dengel, A. (2011). "ICDAR 2011 robust reading competition challenge 2: Reading text in scene images" In Proceedings of the ICDAR pp.1491-1496.

[12] Lohmann S, Ziegler J, Tetzla L "Comparison of tag cloud layouts: task-related performance and visual exploration", in Proceedings of IFIP TC 13 International Conference pp:392-404 (2009)

[13] Arthur, David, and Sergi Vassilvitskii. "K-means++: The Advantages of Careful Seeding." SODA 07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. 2007, pp. 10271035.

[14] J. Matas, O. Chum, M. Urban, and T. Pajdla. "Robust wide baseline stereo from maximally stable extremal regions." Proc. of British Machine Vision Conference, pages 384-396, 2002.

[15] Cortes, C.; Vapnik, V. (1995). "Support-vector networks". Machine Learning. 20 (3): 273297. doi:10.1007/BF00994018

[16] Briandais, Ren (1959). File searching using variable length keys. Proc. Western J. Computer Conf. pp. 295298.