International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

Tamil text detection in videos using Gradient Vector Flow and Fuzzy C-Means

J.T.Anita Rose, B.Jaison, S.V.S.Harshavardhan, B.Krishna Teja, J.Dinesh Sai

Abstract: -In videos, detecting text with multifaceted scenarios is perplexing. Texts in those videos have content full data facts that will be applied for various applications. Here, a system is proposed to enrich the text detection process from video. Here, a new method is implemented that detects Tamil text based on Gradient Vector Flow (GVF) and fuzzy c-means. First the video is split into number of frames. To circumvent temporal redundancy in each frame, a key frame is chosen and the frame where the text be located is identified to be the key frame. The dominant edge pixel is identified in that frame by the sobel edge map. Edge components are detected conforming towards the dominant pixel in sobel detector for constructing Text Candidates (TC).Clustering of a pixel is performed to detect text by using fuzzy c means clustering algorithm. Finally text is detected.

Keywords : Gradient vector flow, Dominant pixel, Text candidate, Fuzzy C-means.

I. INTRODUCTION

Text entrenched in videos encompasseshuge number of valuablecontent that are applied to many real time applications. The semantic content present in the video is an important and useful information to be required for the processing of many real world applications [8,12,25]. Caption text and Graphics text – these are the different kinds of text available while processing with the text present in the video.

Scene texts are those texts that exist in the scenario of shooting the video. The scene text may be for example, the number present on the t-shirt of a football player, the name board of a shop, the license plate details of a vehicle. Scene text is also called graphics text. The Caption text is manually superimposed over the captured video to enhance the meaning of video/image. It is also called superimposed text. Plenty of methods exist for the purpose of extracting the useful information present in images throughout the past decade. This technique includes methods that deal with the texture properties of the image [5], small components of text combined to create words of text [11], edge identifying frameworks [12] and gradient based approaches [21]. Texture based methods extricate text regions along thecontextualcontent [13,16,26].

The approaches in connected component employ the principle of connecting small set of components to form text candidate regions [17, 23]. The approaches employed in edge based apply the principle of finding the boundaries of the image [1, 15, and 24].

The connected component method does not produce good results for images on multifarious backdrop of the image. The texture based methods works superior for multifaceted contextual content in videos compared to connected component approaches. The mixture of edge based method along with gradient feature based method performs better and becomes less apparent in processing of text detection of Tamil language for many reasons. The works in [22] and [28] primarily focuses on Multi-oriented text. In [20] a method based on laplacian was introduced to deal with the multi-oriented text. Our previous work on text detection is in [30, 31, 32, and 33]. As a result, a new-fangled way of methods is the need of the hour to detect Tamil text based on Gradient Vector Flow (GVF) and fuzzy c-means. This method will improve accuracy to detect Tamil text in videos.

II. PROPOSED METHOD

A new method is proposed that detects Tamil text based on Gradient Vector Flow (GVF) and fuzzy c-means. First the video is split into number of frames. To circumvent temporal redundancy in each frame, a key frame is chosen and the frame where the text be located is identified to be the key frame. The dominant edge pixel is identified in that frame by the sobel edge map. Edge components are detected conforming towards the dominant pixel in sobel detector for constructing text Candidates (TC).Clustering of a pixel is performed to detect text by using fuzzy c means clustering algorithm. Finally text is detected.

The Tamil language has different orientation comparable to English language. To tackle Tamil language text, we propose a novel method go detect dominant pixel by using GVF and Fuzzy C-means approach is employed for cluster formation on the pixels and thereafter the process of extracting useful text .

1) *Splitting of frames and key frame selection:*

The video which contains Tamil text is split into number of frames. To circumvent temporal redundancy in each frame, a key frame is chosen and the frame where the text be located is identified to be the key frame.

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

Dr. J.T.Anita Rose*, Associate Professor, Department of CSE, St.Joseph's College of Engineering, Tamil Nadu, India. Email: anitaswartz14@gmail.com

Dr.B.Jaison, Associate Professor, Department of CSE, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India. Email: bjn.cse@rmkec.ac.in

S.V.S.Harshavardhan, Student, IV Year CSE, Department of CSE, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India. Email: harshavardhan1234567891@gmail.com.

B.Krishna Teja, Student, IV Year CSE, Department of CSE, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India. Email: j.dineshsai8358@gmail.com

Figure 1(a) shows the splitting of frames, the results of figure 1 (a) shows that the frame rate is considerably reduced which is the principal cause for the occurrence of temporal redundancy.

The temporal redundancy is circumvented based on comparison between edges of two images that lie close to each other. The sobel edge detection is employed for the purpose of detecting the edge present is a frame. The frame similarity is checked between two frames. Similar frames contain the similarity ratio to be high whereas the dissimilar frames has the similarity ratio as low. The high indexed similarity ratio frames are stored. The low indexed similarity ratio frames are removed.. Figure 1(b) depicts the key frame.

Figure 1(a) :Splitting of videos

Figure 1(b) Key frame selection

2) GVF Process

The key frame is taken as an input to find centraldominant pixel. The Gradient Vector Flow (GVF) is given in equation (1) [19].

$$\alpha = \iint \left(x_i^2 + x_j^2 + y_i^2 + y_j^2\right) + |\nabla f|^2 \, |h - \nabla g^2| didj \quad (1)$$

Where $h(i,j)$ gives the GVF vector field, $g(i,j)$ gives the edgeplot of the input image.

The key frame detected after the splitting of frames undergoes the sobel edge detection phase. Here, the prevailing pixels are discovered for the classifying text pixels against non-text pixels. Figure 2(a) illustrates input image and Figure 2(b) illustrates prevailing pixels of input image.

(a) input **(b) dominant pixel**

Figure 2 (a)input frame (b)dominant pixel selection for the input frame.

Figure 3(a) illustrates GVF for sobel edge and the arrows depict the gradient vector flow. In figure 3 (b), the center pixel exert a pull on four dark shaded arrows and is recognized as the prevailing pixel that contains the text. In figure 3(c), the center pixel exert a pull on only onedark shaded arrow and hence recognized as the pixel to be of non-text.

Figure 3(a): GVF for sobel edge

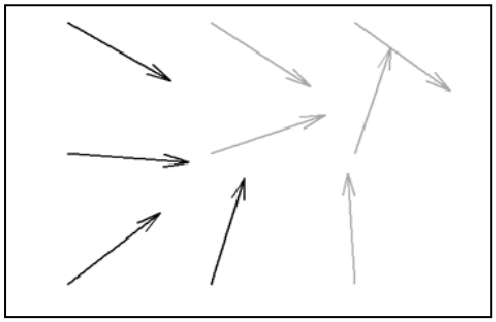


Figure 3(b): GVF arrows illustrating the directions of text pixel

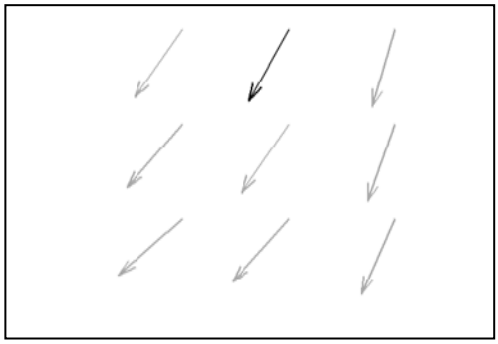


Figure 3(c) GVF arrows illustrating the directions of non-text pixel

3) Text Candidates Formation

The outcome of the process involving the identification of the pixels which are dominant is then followed by the Candidate Text formation. This method extracts edge components present in pixels that are dominant using the conventional Sobel detector. The possible pixels that are identified as edges becomes the possibility of text. Figure 4 shows the text candidate selection. The result of the process detects almost all Tamil text components except with very negligible false positives. Next by Sobel edge map, the text candidates extracted are used to restore the full text information.

Figure 4: Text candidate selection

4) Clustering of text candidates using different color layers.

From the text candidate generated, the frame is grouped to identify text pixel from non text pixel. The grouping is performed with different color layers and gray scale components. The main aim of this step is to form a word by merging segments of character components and neighbor characters. The color informationpresentin text is diverse from those present in the background. There is also some false positive will remain in thisphase. To eliminate false positive skeletonization is done tocheck for false positive figure 6 shows the color layers without false positive.

5) Skeletonization

Figure 5 depicts the skeleton structure of frame.

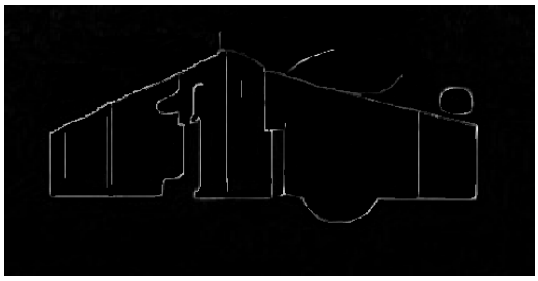


Figure 5:Skeleton structure of frame.

Figure 6: Grouping of color components of text

6) Text detection using fuzzy C-means clustering

Output from figure 6 gives separate color for text and its background. Fuzzy c-means clustering groups text region separately from non-text region. This allows one Square pixel to belong to two or more clusters [3, 7].

The function Q_t , the cluster formation function using fuzzy c-means is given in equation (4).

$$Q_t = \sum_{k=1}^{N} \sum_{l=1}^{Q} f_{kl}^{t} \left| X_k - O_l \right|^2 , 1 \le t < \infty \quad (4)$$

Where,

t is a number that is bigger than 1,

u_{kl} is membership degree,

x_k is the k^{th} data,

O_l is cluster center.

Partitioning of fuzzy is shown in equation (5), first of all initialize $F = f_{kl}$,$F^{(0)}$ then calculate the center at m steps by equation (6) and update F^{m}, F^{m+1} by equation (5)

With updation of member f_{kl} and the cluster centers o_l by,

$$f_{kl} = \frac{1}{\sum_{m=1}^{O} \left(\frac{|X_k - o_l|}{|X_k - o_m|}\right)^{\frac{2}{m-1}}} \quad (5)$$

Where, o_l is

$$o_l = \frac{\sum_{k=1}^{N} f_{kl}^{m} \cdot X_k}{\sum_{k=1}^{N} f_{kl}^{m}} \quad (6)$$

This above iteration will stop when $max_{kl}\left\{\left|f_{kl}^{(n+1)} - f_{kl}^{(n)}\right|\right\} < \varepsilon$ and ε is the termination point and it is between zero and one and n is the number of iterated steps. The iteration procedure will stop when it come to minimal point of Q_t[3,7]. Finally text is detected by this fuzzy c-means algorithm, but still there is some false positives remaining.

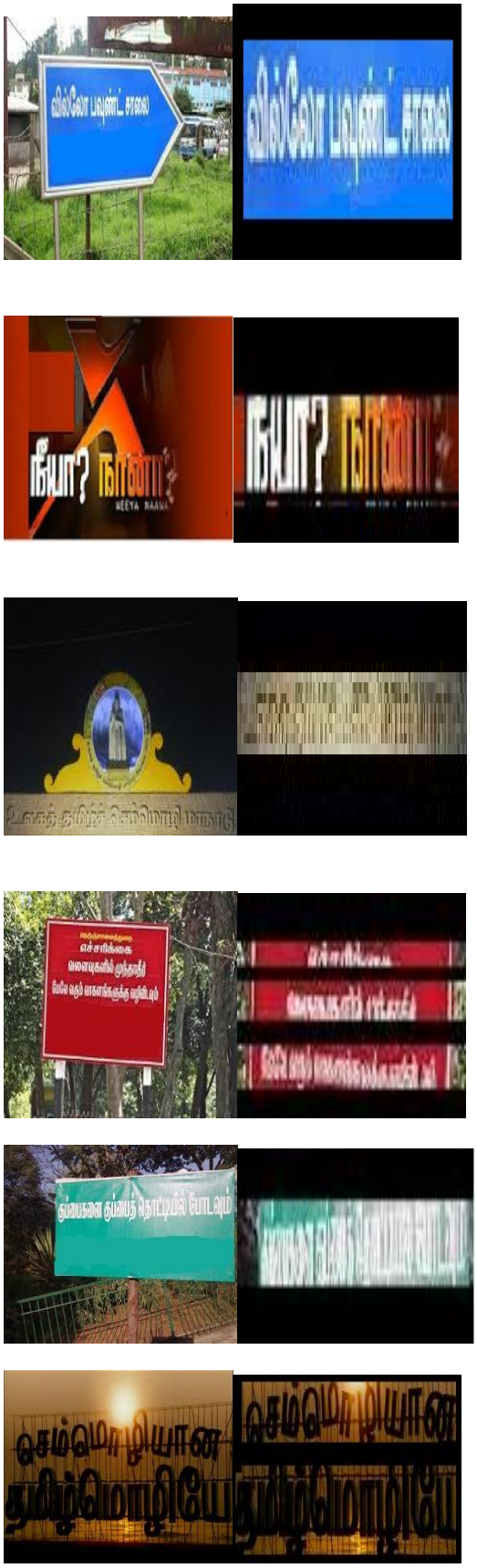
Figure 7: Text detected by fuzzy c-means

III. EXPERIMENTAL RESULTS:

The proposed system is tested and evaluated with our new dataset. Our dataset consists of videos taken from news videos, advertisement videos, TV shows videos and videos from google that contain the text information. The dataset contains 2500 videos. The proposed system was evaluated for Precision and recall. Table 1 gives the experimentation results. Our system produced 79% precision and 89% recall rate.

Table 1 Experimental evaluation

	Precision	Recall
Our proposed method	79%	89%
Becker H	62%	67%
Chen A	60%	60%
Ashhida	55%	46%
David H W	44%	46%
Zhu Q	33%	40%
Wolf	30%	44%
J.Kim	22%	28%
Todoran	19%	18%
N.Ezaki	18%	36%



(a) (b)

(a) (b)

(a) **(b)**

Figure 8(a) Original input (b) text detected output

IV. CONCLUSION

A new method to detect text for Tamil language is proposed using Gradient Vector Flow (GVF) and fuzzy c-means.Dominant edge pixel is identified in that frame by the sobel edge map.Edge components are detected using sobel edge map. Clustering of a pixel is performed to detect text by using fuzzy c means clustering algorithm. Finally text is detected.

REFERENCES

1. A, Alshennawy and A, Aly, (2009), 'Edge detection in digital images using fuzzy logic technique', World Academy of Science, Engineering and Technology .
2. Ashida, H, KurikiI, Murakami, I, Hisakata ,R and Kitaoka, A, (2012),'Direction-specific fMRI adaptation reveals the visual cortical network underlying the Rotating Snakes illusion'.
3. J, C, Bezdek, (1981), 'Pattern Recognition with Fuzzy Objective Function Algoritms', Plenum Press, New York.
4. F ,Chassaing , C, Wolf, J, M,Jolion, (2002), 'Text localization, enhancement and binarization in multimedia documents,' International Conference onPattern Recognition, vol. 2, pp. 1037–1040.
5. D, Chen, J,M ,Odobez and J,P,Thiran,(2004), 'A localization/verification scheme for finding text in images and video frames based on contrast independent features and machine learning', Signal

6. Processing image communication, pp. 205-217.
7. David Doermann,Jian Liang, Huiping Li, (2003), 'Progress in camera-based document image analysis', International Conference on Document Analysis and Recognition, pp. 606-616.
8. J C Dunn, (1973), 'A Fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters', Journal of Cybernetics, vol. 3, pp.32-57.
9. B, Epshtein, E ,ofek and Y, Wexler,(2010), 'Detecting text in natural scenes with stroke width transform', CVPR, pp. 2963-2970.
10. N ,Ezaki, M ,Bulacu and L, Schomaker, (2004),'Text detection from natural scene images: Towards a system for visually impaired persons', International Conference on Document Analysis and Recognition, pp. 683-686.
11. X S Hua, L Wenyin and HJ Zhang,(2004), 'An automatic performance evaluation protocol for video text detection algorithms', IEEETransactions on CSVT, pp. 498-507.
12. A Jamil, I Siddiqi, F Arif and A Raza, (2011), 'Edge-based features for localization of artificial urdu text in video images',International Conference on Document Analysis and Recognition, pp. 1120-1124.
13. K Jung, K I Kim, and A K Jain, (2004), 'Text information extraction in images and video: A survey,' Pattern Recognition, vol. 37, pp. 977–997.
14. K Kim, KJung, and J Kim, (2003), 'Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm,' IEEE Transactions on Pattern Analysis Machine Intelligence, vol. 25, pp. 1631–1639.
15. L,Todoran, M, Aiello, M, Worring, and C,Monz, (2001),'Document understanding for a broad class of documents', Technical Report Intelligent Sensory Information Systems Group, University of Amsterdam.
16. M R Lyu, J Song, and M Cai, (2005), 'A comprehensive method for multilingual video text detection, localization, and extraction,' IEEE Transactions on circuits systemsvideo technology, vol. 15, pp.243–255.
17. W Mao, F Chung, K K M Lam, and W Sun, (2002), 'Hybrid Chinese/English text detection in images and video frames,' International Conference onPattern Recognition, vol. 3, pp. 1015–1018.
18. VY Mariano and R Kasturi, (2000), 'Locating uniform-colored text in video frames,'International Conference onPattern Recognition, pp. 539-542.
19. PalaiahnakoteShivakumara, TrungQuy Phan, Shijian Lu and Chew Lim Tan, (2013), 'Gradient Vector Flow and Grouping based Method for Arbitrarily-Oriented Scene text Detection in Video Images' ,IEEE transactions on circuits and systems on video technology , pp. 1729-1739.
20. TRodt, HPBurmeister, SBrtling, J Karminsky, BSchwab, RKikinis, HBecker, (2004),'3D visualisation of the middle ear by computer assistant post-processing of Helicle multi slice CT data' pp.1684-1692.
21. P, Shivakumara, T, Q, Phan and C, L, Tan, (2011), 'A Laplacian Approach to Multi-Oriented text detection in video', IEEE Transactions on PAMI, pp.412-419.
22. [21] X, Tang, X, Gao, J, Liu, and H, Zhang,(2002), 'A spatial-temporal approach for video caption detection and recognition,' IEEE transactions on neural networks, vol. 13, no. 4, pp. 961–971.
23. H Tran, A Lux, H L T Nguyen and A Boucher,(2005), 'A novel approach for text detection in images using structural features', ICAPR, pp. 627-635.
24. E,K, Wong and M, Chen, (2003), 'A new robust algorithm for video text extraction', Pattern Recognition, vol. 36, pp. 1397-1406.
25. C, Xu and J, L, Prince, (1998), 'Snakes, shapes, and Gradient Vector Flow', IEEE transactions on image processing, pp. 359–369.
26. YasarBecerikli, Tayfun M Karan and Ali Okatan, (2009), 'A new Fuzzy edge detector for noisy images using modified WFM filter', International journal of innovative computing, information and control, vol. 5.
27. J Zang and R Kasturi, (2008), 'Extraction of text objects in video documents: recent progress', DAS, pp. 5-17.
28. Y Zhong, H Zhang, and A K Jain,(2000), 'Automatic caption localization in compressed video,' IEEE Transactions on pattern analysis andmachine Intelligence, vol. 22, no. 4, pp. 385–392.
29. J Zhou, L Xu, B Xiao and R Dai, (2007), 'A robust system for text extraction in video', ICMV, pp.119-124.
30. [29] Zhu, Q Yeh, MC Cheng, (2006),'Multimodal fusion using learned text concepts for image categorization', ACM International conference on Multimedia', pp. 211–220.
31. A Thilagavathy, K Aarthi, A Chilambuchelvan, (2012), 'A hybrid approach to extract scene text from videos' , 'International conference on computing electronics and electrical technologies, pp. 1017-1022.
32. ThilagavathyA , AarthiK, Chilambuchelvan A, (2012) 'Text Extraction from videos using a hybrid approach', 'International conference on advances in computing, communications and informatics ' , pp. 193 – 199.
33. ThilagavathyA , Aarthi K, Chilambuchelvan A, (2012), 'Text detection and extraction from videos using ANN Based Network', 'International journal on soft computing, artificial intelligence and applications' , vol 1, pp: 19-28.
34. Thilagavathy A &Chilambuchelvan A, (2017). 'Fuzzy based edge enhanced text detection algorithm using MSER', Cluster computing. 10.1007/s10586-017-1448-5.