An Efficient Approach for Text Extraction in Images and Video Frames Using Gabor Filter

Anubhav Kumar, *Member, IACSIT*

Abstract—**In this paper, an effective methodology for text extraction images and video frames using Gabor filter is proposed. The proposed approach is completed by Gabor Filter, morphological and Heuristic filtering process methods is used to localize the text region better. The proposed technique is completed by text extraction utilizing Gabor filter method which is utilized for text identification within complex images and video frames. Diverse experiments were led to assess the execution of the proposed calculation and algorithm and compare with other methods. Experimental results tested from a large dataset and demonstrated that the proposed method is effective and practical. Various parameters like a precision and recall rates are analyzed for both existing and proposed method to determine the success and limitation of our method. Experiment results show that our method can obtain 99.11 % recall rate and precision rate 94.67 % with average computational time 5.28 second /frames.**

Index Terms—**Text extraction, text localization, text recognition, gabor filter.**

I. INTRODUCTION

Text extraction from images remains a challenging issue for image processing applications, due to complex background, unknown text color, and different language characteristics. Robust detection of text from multimedia, document and internet is a challenging problem. A few methods have been proposed for text detection and detection in images or video frames. The proposed algorithm is powerful for text lines of all font sizes and styles, as long as they are excessively small little or huge in respect to the image frame.

Generally, Text extraction, detection and recognition methods can be classified into three categories, texture-based, connected component based and edge-based. The texture based methods analysis, based on Gabor filtering and wavelet analysis. All these methods are quite general and flexible but they are also computationally demanding. Shivakumara *et al.* [1] based on texture methods and proposed a new method for text detection in images based on combination of wavelet and color features for text detection in video. Yan *et al.* [2] Gabor filters with scale and direction varied to describe the strokes of Chinese characters for candidate text area extraction. Pati *et al.* [3] Proposed a biologically inspired, multichannel filtering scheme for page layout analysis using Gabor filter. Yan Gllavata *et al.* [4] proposed an efficient algorithm which can automatically detect, localize and extract horizontally aligned text in images using connected component methods. Connected component based methods [4] segment an image into a set of connected components and successively merge the small components into larger ones. Kumar *et al.* proposed edge based methods [5], [6] an efficient text extraction algorithm in complex images and an efficient algorithm for text Localization and extraction in complex video text images. A focus of line detection mask based system for text region localization has been proposed by Liu *et al.* [7]. A. Kumar *et al.* [8], [9] proposed a line detection mask based text extraction in images and video frames. Edge based methods based on detection masks and localized image using heuristics filtering with OCR recognition.

W. Kim and C. Kim [10] proposed method is robust to different character size, position, contrast, and color. It is also language independent. Overlay text region update between frames is also employed to reduce the processing time.

A robust system is proposed by V. Wu [11] to automatically detect and extract text in images from different sources, including video, newspapers, advertisements, stock certificates, photographs, and checks. X. Gao *et al.* [12] present algorithms for detection, extraction, binarization and recognition of Chinese video captions. Q. Ye *et al.* [13] proposed a novel coarse-to-fine algorithm that is able to locate text lines even under complex background by using multiscale wavelet features.

K. Kim *et al.* [14] proposed texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. X. J. Li *et al.* [15] presents a fast and effective approach to locate text lines even under complex background. K. C. Kim *et al.* [16] proposed a method that extracts text regions in natural scene images using low-level image features and that verifies the extracted regions through a high-level text stroke feature. C. Wolf *et al.* [17] present an algorithm to localize artificial text in images and videos using a measure of accumulated gradients and morphological post processing to detect the text.

In this paper an efficient approach proposed for text extraction in images and video frames based on Gabor filter, Text area localization and text recognition steps.
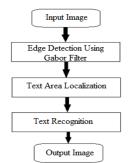

Fig. 1. Text image extraction block diagram.

Our proposed method an efficient methodology for content and text extraction in images and video frames using Gabor filter separates a text area from an image. Essentially it works in three steps described in Fig. 1. The rest of the paper is organized as follows. Section II is described proposed algorithm and the result and analysis is done in Section III. The conclusion process in Section IV has been done.

II. PROPOSED ALGORITHM

Prior to describing the details, proposed method text extraction method is outlined here with the help of the flow chart in Fig. 2. Basically, the proposed Gabor filter based text extraction method comprises three steps: edge detection using Gabor Filter, Text area localization and text recognition. The Steps of proposed algorithm an edge based method for Text extraction in complex images and video frames using Gabor filter described as follows:
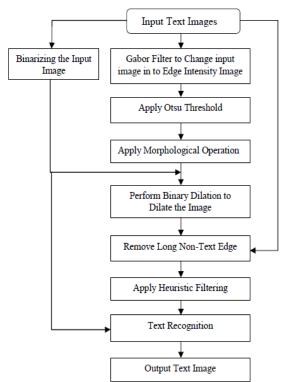

Fig. 2. Flow chart of the proposed algorithm.

A. Edge Detection Using Gabor Filter

In the proposed method, the Gabor Filter Transform property of edge is used to find the edge map from every images and video frames in various orientations. Gabor filters can be sensitive to stripes of a specified width and orientation. The family of two-dimensional Gabor filters $G_{(\lambda, \theta, \varphi)}(x, y)$, which was proposed by Daugman [18], are often used to obtain the spatial frequency of the local pattern in an image.

$$G_{(\lambda,\theta,\Phi)}(x, y) = e^{-\frac{(x'^2+\gamma^2 y'^2)}{2\sigma^2}} \cos(2\Pi x' \frac{1}{\lambda} + \Phi) \quad (1)$$

$$x' = x\cos\theta + y\sin\theta \quad (2)$$

$$y' = -x\sin\theta + y\cos\theta \quad (3)$$

where the arguments x and y represent the pixel coordinators, θ parameter define the spatial aspect ratio and $1/\lambda$ is called the spatial frequency, and specifies the orientation of the filter [19]. Fig. 3 shows that Impulse response of an even-symmetric gabor function in orientation 90°.

Fig. 3. Impulse response of an even-symmetric gabor function in orientation 90°.

These filters provide the optimal resolution for both the orientation and the spatial frequency of a local image region and with this help find out edge in images. So Gabor-based asymmetric filters is introduced that is edge-form filters used to obtain the precise scale information of the located edges in an image. The edge-form filters $E_{(\lambda, \theta)}(x, y)$, are the Gabor filters with $\Phi = \Pi / 2$ [19].

$$E_{(\lambda,\theta)}(x, y) = e^{-\frac{(x'^2+\gamma^2 y'^2)}{2\sigma^2}} \cos(2\Pi x' \frac{1}{\lambda} + \frac{\Pi}{2}) \quad (4)$$

$$x' = x\cos\theta + y\sin\theta \quad (5)$$

$$y' = -x\sin\theta + y\cos\theta \quad (6)$$

So, the original image should be converted into edge image by Gabor filter transform. The conversion result is shown in Fig. 4. After Gabor filter, firstly contrast is increased by sharpening filter in the proposed method. After that Otsu thresholding has been done to find the binary image. The Otsu method is a histogram-based global thresholding method. Therefore, it can extract text pixels from simple background and change pixel contrast.

Fig. 4. (a) Original input image (b) edge based intensity image using Gabor filter.

B. Text Localization Using Heuristic Filtering

In the localization process, firstly morphologically open binary process is used to convert the binary image to dilate image. The main purpose of dilation operation on a binary image is basically enhancing the boundaries of area of foreground pixels by image boundaries pixels of objects.

After that the non-text long edge is found by applying the multiplication process in dilated and input binary image. Now, the long edge of final refined image is removed with the help of connected component labeling operator and 4-neighbour connected component are utilized. Finally Heuristic Filtering supports to find out non-text region by using major to minor axis ratio. Basically those regions removed which have width/height less than 1/10 and an area greater than or equal to 1/20 of the maximum area region [6]. After this morphological operation is used to automatically take advantage of the decomposition of a structuring element object. Also, when performing binary dilation with a structuring element object that has a decomposition, imerode automatically uses binary image packing to speed up the dilation Retained image shown in Fig. 5.

Fig. 5. Text area localization output.

C. Text Recognition

In text recognition, the common OCR systems is available requires the input image to be such that the characters can be easily parsed and recognized. After that the Text localization images multiply with the input binary image. At last this process generates an output image with white text against a black background. Final text recognition output is shown in Fig. 6.

Fig. 6. Text extraction result.

III. RESULT AND ANALYSIS

Experiments have been carried out on large set of camera captured images varying with respect to size color and orientation. We tested our text extraction procedure on 32 complex images and video frames which are randomly selected from internet randomly having 1578 total text words. On our personal laptop of Pentium 4 the experiment has been carried out at 2.4 GHz.

From Fig. 7-Fig. 11, we can see that the performance of our proposed technique and other existing methods on a wide assortment of image and video frames. The execution of every procedure has been assessed dependent upon its precision and recall rates obtained overall.

$$Recall\ Rate = (Correctly\ detected\ words/correctly\ detected\ words + False\ Negative) \times 100 \quad (7)$$

$$Precision\ Rate = (Correctly\ detected\ words\ /correctly\ detected\ words + False\ positive) \times 100 \quad (8)$$

Fig. 7. (a) Original image (b) proposed method output (c) Liu [6] method output (d) gllavata [4] method output.

Fig. 8. (a) Original image (b) Proposed method output (c) Liu [6] method output (d) Gllavata [4] method output.
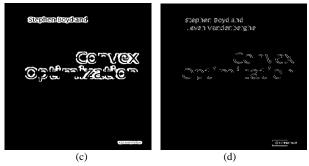
(c)   (d)

Fig. 9. (a) Original image (b) Proposed method output (c) Liu [6] method output (d) Gllavata [4] method output.

From Table I and Table II, results obtained shows the cases of complex images and video frames, the average precision (94.22%) and recall (99.44%) rates obtained from the Indoor, Outdoor and Books cover Images and the average precision (95.05%) and recall (98.33%) rates obtained from the News, Commercial and Video Frames. In all the images recall rate of 99.11, precession rate of 94.67 and overall computed time of 5.28 second can be achieved by the experiment results of the proposed detection method.

TABLE I: RESULT OF THE PROPOSED METHOD

Image and Video Frames type	Proposed Method		
	Total words	Detected words	False positive
Indoor, Outdoor and Books cover Images	722	718	44
News, Commercial and Video Frames	856	846	44

TABLE II: RESULT OF THE PROPOSED METHOD

Image and Video Frames type	Proposed Method		
	Recall Rate %	Precision Rate %	Average Time f/s
Indoor ,Outdoor and Books cover Images	99.44	94.22	7.01
News ,Commercial and Video Frames	98.83	95.05	3.81
Total Results	99.11	94.67	5.28

(c)   (d)

Fig. 10. (a) Original image (b) proposed method output (c) Liu [6] method output (d) Gllavata [4] method output.

(a)   (b)

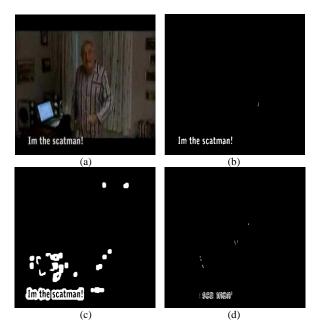(c)   (d)

Fig. 11. (a) Original image (b) proposed method output (c) Liu [6] method output (d) Gllavata [4] method output.

TABLE III: RESULT OF THE PROPOSED

Methods	Recall rate %	Precision Rate %
Proposed	99.11	94.67
Gllavata *et al* [4]	88.7	83.9
Liu *et al*.[7]	96.6	91.8
Ye *et al* [13]	90.8%	-
Kim *et al*.[14]	82.8	63.7
Li *et al*[15]	91.1%	-
Wolf *et al*.[17]	93.5	-

Table III shows the performance comparison of our proposed method with other existing text extraction and detection method, where our proposed method has a better performance in precision rate and recall rate. The reason for better recall rate that the proposed method used Gabor filter to change image in to edge based image.

IV. CONCLUSION

In this paper, an efficient approach is used for text extraction in images and video frames using Gabor filter. An experimental result shows that our proposed method is very effective and efficient in detection and extraction text from different images and video frames. Experiment results show that our method can obtain 99.11 % recall rate and precision rate 94.67. The algorithm works reasonably well given the complexity of the complex images and video frames, suggesting that such techniques could prove useful in information retrieval applications in images and video frames.

REFERENCES

[1] P. Shivakumara, T. Q. Phan, and C. L. Tan, "New wavelet and color features for text detection in video," in *Proc. 20th International Conference on Pattern Recognition*, 2010, pp. 3996–3999.

[2] J. Q. Yan, D. C. Tao, C. Tian, X. B. Gao, and X. L. Li, "Chinese text detection and location for images in multimedia messaging service," in *Proc. IEEE International Conference on Systems Man and Cybernetics*, 2010, pp. 3896–3901.

[3] P. B. Pati, S. Raju, N. Pati, and A. G. Ramakrishnan, "Gabor filters for document analysis in Indian bilingual documents," in *Proc. International Conference on Intelligent Sensing and Information Processing*, 2004, pp. 123–126.

[4] J. Gllavata, R. Ewerth, and B. Freisleben, "A robust algorithm for text detection in images," in *Proc. the 3rd International Symposium on Image and Signal Processing and Analysis*, 2003, vol. 2, pp. 611–616.

[5] A. Kumar, "An efficient text extraction algorithm in complex images," in *Proc. Sixth International Conference on Contemporary Computing*, 2013, pp. 6-12.

[6] A. Kumar and N. Awasthi, "An efficient algorithm for text Localization and extraction in complex video text images," in *Proc. 2nd International Conference on Information Management in the Knowledge Economy*, 2013, pp.14-19.

[7] X. Q. Liu and J. Samarabandu, "Multiscale edge-based text extraction from complex images," in *Proc. International Conference on Multimedia and Expo*, 2006, pp. 1721-1724.

[8] A. Kumar, A. K. Kaushik, R. L. Yadava, and D. Saxena, "An edge-based algorithm for text extraction in images and video frame," *Advanced Materials Research*, vol. 403-408, pp. 900-907, 2012.

[9] A. Kumar, A. K. Kaushik, R. L. Yadav, and Anuradha, "A robust and fast text extraction in images and video frames," in *Proc. the Springer International Conference of Advances in Computing, Communication and Control*, 2011, pp. 342-348.

[10] W. Kim and C. Kim, "A new approach for overlay text detection and extraction from complex video scene," *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 401–411, 2009.

[11] V. Wu, R. Manmatha, and E. M. Riseman, "Textfinder: an automatic system to detect and recognize text in images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 11, pp. 1224–1229, Nov. 1999.

[12] X. Gao *et al.*, "Automatic news video caption extraction and recognition," in *Proc. LNCS 1983: 2nd Int. Conf. Intell. Data Eng. Automated Learning Data Mining, Financial Eng., Intell. Agents*, Hong Kong, 2000, pp. 425–430.

[13] Q. Ye, Q. Huang, W. Gao, and D. Zhan, "Fast and robust text detection in images and video frames," *Image and Vision Computing*, vol. 23, pp. 565-576, 2005.

[14] K. Kim, K. Jung, and J. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1631–1639, Dec. 2003.

[15] X. J. Li, W. Q. Wang, S. Q. Jiang, Q. M. Huang, and W. Gao, "Fast and effective text detection," in *Proc. 15th IEEE International Conference on Image Processing*, 2008, pp. 969–972.

[16] K. C. Kim, H. R. Byun, Y. J. Song, Y. M. Choi, S. Y. Chi, K. K. Kim, and Y. K. Chung, "Scene text extraction in natural scene images using hierarchical feature combining and verification," in *Proc. the 17th International Conference in Pattern Recognition*, 2004, vol. 2, pp. 679–682.

[17] C. Wolf, J. M. Jolion, and F. Chassaing, "Text localization, enhancement and binarization in multimedia documents," in *Proc. 16th International Conference on Pattern Recognition*, 2002, vol. 2, pp. 1037–1040.

[18] G. Daugman, "Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America A*, vol. 2, issue 7, pp. 1160-1169, 1985.

[19] D. Chen, K. Shearer, and H. Bourlard, *11th International Conference on Image Analysis and Processing Proceedings*, pp. 192–197, 2001.

Anubhav Kumar was born in Bulandshahr, Uttar Pradesh, India on March 30, 1986. He received his B.Tech degree in electronics and communication engineering at Uttar Pradesh Technical University and M.Tech degree in electronics and communication engineering (specialization in image processing) at Uttar Pradesh Technical University. He was with Vishveshwarya Institute of Engineering and Technology, G. B. Nagar (U.P), India. He is now an assistant professor with the Department of Electronics and Communication, Raj Kumar Goel Institute of Technology for Women, Ghaziabad, India. He is a member of IEEE, IACSIT and many international societies. He also has many years' research and academic experience in signal & image processing and microwave engineering.

Networking and Navigation