International Journal of Trend in Research and Development, Volume 3(2), ISSN: 2394-9333
www.ijtrd.com

Arbitrarily Oriented Scene Text Detection in Video Images using Multi-Spectral Fusion Based Approach

P.Santosh kumar[1], M.Balasubramanyam[2] and B.Suneetha[3]
[1.]M.tech student, ECE department, RISE Krishna Sai Prakasam Group of Institutions, Ongole, India
[2]Assistant Professor, ECE department, RISE Krishna Sai Prakasam Group of Institutions, Ongole, India
[3]Sr.Assistant professor, ECE department, Dadi Institute of Engineering & Technology,Visakahapatnam,India

Abstract: Scene text detection from video as well as natural scene images is challenging due to the variations in background, contrast, text type, font type, font size, and so on. Besides, arbitrary orientations of texts with multi-scripts add more complexity to the problem. The proposed approach introduces a new idea of convolving Laplacian with wavelet sub-bands at different levels in the frequency domain for enhancing low resolution text pixels. Then, the results obtained from different sub-bands (spectral) are fused for detecting candidate text pixels. We explore maxima stable extreme regions along with stroke width transform for detecting candidate text regions. Text alignment is done based on the distance between the nearest neighbor clusters of candidate text regions. In addition, the approach presents a new symmetry driven nearest neighbor for restoring full text lines. We conduct experiments on our collected video data as well as several benchmark data sets, such as ICDAR 2011, ICDAR 2013, and MSRA-TD500 to evaluate the proposed method. The proposed approach is compared with the state-of-the-art methods to show its superiority to the existing methods.

Keywords: *Laplacian-Wavelet, Multi Spectral Fusion, Maxima Stable Extreme Regions, Stroke Width Transform, Arbitrarily Oriented Video Text Detection.*

I. INRODUCTION

With the recent advances in multimedia and network technologies combined with the rapid decline in hardware prices, the contents of digital video and images are growing at a tremendous speed. As the statistics data of 2010 shows, more than 35 hours of video contents were uploaded to video sharing sites (e.g., YouTube) every minute and more than 35 million photos had been uploaded to social networking sites like Facebook every month [1]. This results in huge databases and thus requires approaches which work at high level semantics. The conventional approaches that use low level features may not be sufficient for handling such large databases due to the gaps between low level features and high level semantics. To alleviate this problem, text detection and recognition has become popular as it provides meaningful cues which are close to the content of video or image [2]–[4]. So, it has been widely used in video summarization, content based image indexing and video sequence retrieval. On top of these applications, text detection and recognition has also been used for real time surveillance applications, such as assisting a blind person to walk freely on roads, assisting tourists to reach their destinations, enhancing safe driving, navigating vehicles based on license plate information, exciting event extraction from sports video, identifying athletes in marathon events, etc [5]. Video consists of two types of texts, namely, caption text and scene text. Caption text is manually edited, which has good clarity and visibility and hence is easy to process. Scene text exists naturally in video frames, the detection of which suffers from color bleeding, low contrasts, low quality due to distortion, different orientations, backgrounds, etc. Hence, scene text is hard to process compared to caption text [4], [6], [7]. Scene images captured through a high resolution camera usually contain only scene texts with high contrast and complex background, while video contains both caption and scene texts with low resolution and complex background. Achieving good accuracy for text detection from both video and natural scene images is still an open issue in the field of image processing and pattern recognition because most of the existing approaches [8], [9] either focus on caption text in video or scene text in natural scene images but not both video and natural images. The problem of text detection and recognition from scanned document images is not new for the document analysis community because for different scripts we can find several Optical Character Recognizers (OCR engines) that are available publicly. However, the same methods may not be used for detection and recognition of the texts in video and natural scene images because the approaches work well for plane background and high contrast images but not for images like video and natural scene images [2], [10], [11]. Text Extraction from images is a major task in computer vision. Applications of this task are various (automatic image indexing, visual impaired people assistance or optical character reading...). Many studies focus on text detection and localization in images. However, most of them are specific to a constrained context such as automatic localization of postal addresses on envelopes [1], license plate localization [2], text extraction in video sequences [3], automatic forms reading [4] and more generally ”documents” [5]. In spite of such extensive studies, it is still not easy to design a general-purpose TIE system [6]. This is because there are so many possible sources of variation when extracting text from a shaded or textured background, from low-contrast or complex images, or from images having variations in font size, style, color, orientation, and alignment. These variations make the problem of automatic TIE extremely difficult. Increasing popularity of digital cameras and camera phones enables acquisition of image and video materials containing scene text, but these devices also introduce new imaging conditions such as sensor noise, viewing angle, blur, variable illumination etc. Taking into account all these problems and scene text properties it is clear that its extraction and recognition is more difficult task in comparison with caption text and text in documents. Text information extraction consists of 5 steps [7]: detection, localization, tracking, extraction and enhancement, and recognition (OCR). In case of scene text particular focus is set on extraction. This step is done on previously located text area of image and its purpose is segmentation of characters from background that is separation of text pixels from background pixels. Text extraction strongly affects recognition results and thus it is important factor for good performance of the whole process. Text extraction methods are classified as threshold based and grouping-based. First category includes histogram-based thresholding [8], adaptive or local thresholding [9] and entropy-based methods. Second category encompasses clustering-based, region based and learning-based methods. Clustering techniques performed well on color text extraction [10]. Region-based approaches, including region-growing and split and merge algorithm, exploit spatial information to group character pixels more efficiently, but drawback is dependence on parameter values. Learning-based methods mostly refer to multi-layer perceptrons and self-organizing maps, but variation of scene text makes difficult to create representative training database. The proposed work will introduce novel text extraction techniques with Discrete Wavelet Transform and k-Means Clustering. The system also introduces morphological operation like dilation and erosion for segregation of text and non-text regions for better accuracy.

The proposed system as shown in Figure 1 presents a research methodology where the text extraction from images with different scenario deploying discrete wavelet transform and k-means clustering. The prominent edges captured from the input binarized image are estimated using two dimensional discrete wavelet transform. Finally, when this stage is accomplished, morphological operations like erosionand dilation is implemented for the purpose of removing some non-text area which can be easily confused as text region. The morphological operations also associated various segregated candidate text regions in each information for sub-band of the binarized image. The fact in this stage for consideration is that binary information about the colors actually do not assist in text extraction procedure from the given image. The proposed system accepts input as colored RGB image for more real-time environment in development. The image is then processed in wavelet domain and then the text extraction process is implemented in later stage of processing. The proposed discrete wavelet transform system can be exhibited by following flow:

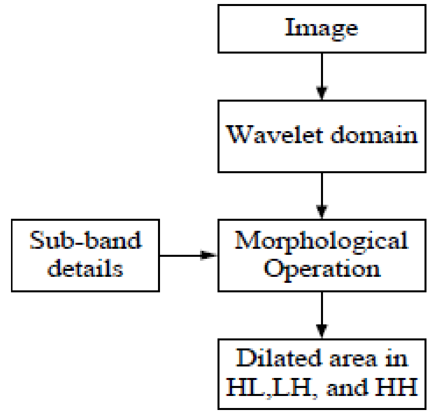

Figure 1: Proposed wavelet based text extraction protocol

II. WAVELET TRANSFORM

Wavelets are functions defined over a finite interval and having an average value of zero.The basic idea of the wavelet transform is to represent any arbitrary function (t) as a superposition of a set of such wavelets or basis functions. These basis functions or baby wavelets are obtained from a single prototype wavelet called the mother wavelet, by dilations or contractions (scaling) and translations (shifts). The Discrete Wavelet Transform of a finite length signal x(n) having N components, for example, is expressed by an N x N matrix. Wavelets are mathematical functions that were developed by scientists working in several different fields for the purpose of sorting data by frequency. Translated data can then be sorted at a resolution which matches its scale. Studying data at different levels allows for the development of a more complete picture. Both small features and large features are discernable because they are studied separately. Unlike the Discrete Cosine Transform, the wavelet transform is not Fourier-based and therefore wavelets do a better job of handling discontinuities in data. In this section we would be employing Haar wavelet transform for image compression. The Haar wavelet operates on data by calculating the sums and differences of adjacent elements. The Haar wavelet operates first on adjacent horizontal elements and then on adjacent vertical elements.

A. Discrete Wavelet Transform

Digital image processing has witnessed a discrete wavelet transform as a prime tool in the area of multi-resolution analysis [21]. 1-D discrete wavelet transform decomposes an input image into mean constituent and detail constituent by estimation with the help of high-pass filter and low-pass filter [22]. Whereas 2D discrete wavelet transform will decompose an input image into 4 sub-bands (LL (*mean constituent*), LH, HL, and HH (*detailed constituent*)).

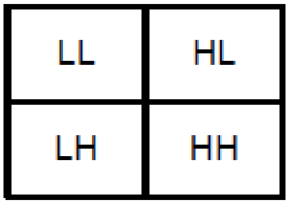

Figure 2: D DWT decomposition output representation

The multi-resolution of the two dimensional wavelet domains can be deployed to explore the text regions of an input image. The conventional filters and detection mechanism for regions can also be expected to provide the equivalent output too. In comparison to one dimensional, 2D discrete wavelet transform can be the better option as it can identify maximum number of edges in one time which cannot be done by conventional algorithms. The conventional boundary detection filters can identify 3 types of boundaries using different types of masking operators as shown in Fig 3. This is also one of the significant reasons of why the conventional boundary detection filters are not faster in comparison to two dimensional discrete wavelet transform.

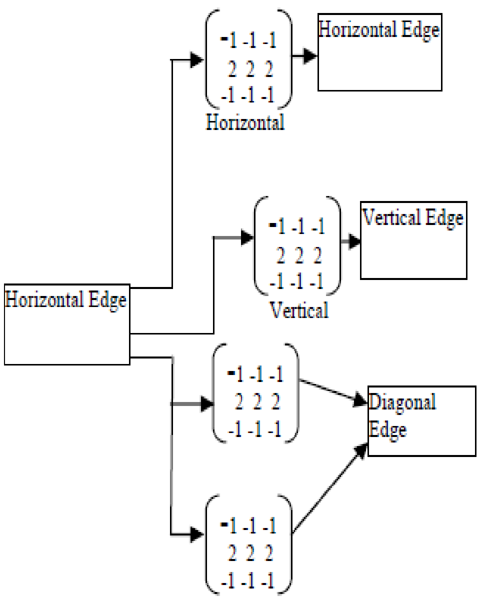

Figure 3: Conventional boundary detection by mask operator

Figure 4: (a) Actual grey image (b) DWT coefficients

A grey scale image when achieved from the original input of RGB image is as shown in Fig 4(a). Fig 4(b) shows how the discrete wavelet transform converts the gray scale image into four sub-bands. The similar operation when performed by Haar [23] discrete wavelet transform makes the processing less complicated, faster with good accuracy, and efficient in comparison to other types of wavelet domain. The important features of the Haar wavelets are very contributing factors in the proposed methodology. The Haar DWT is genuine, symmetric, and orthogonal with simplest boundary situation along with support for random spatial grid distance. It also supports simple high-pass and low-pass filter coefficient [23].

$$\begin{bmatrix} A & B & C & D \\ E & F & G & H \\ I & J & K & L \\ M & N & O & P \end{bmatrix} \qquad \begin{bmatrix} (A+B) & (C+D) & (A-B) & (C-D) \\ (E+F) & (G+H) & (E-F) & (G-H) \\ (I+J) & (K+L) & (I-J) & (K-L) \\ (M+N) & (O+P) & (M-N) & (O-P) \end{bmatrix}$$

(b)

$$\begin{bmatrix} (A+B)+(E+F) & (C+D)+(G+H) & (A-B)+(E-F) & (C-D)+(G-H) \\ (I+J)+(M+N) & (K+L)+(O+P) & (I-J)+(M-N) & (K-L)+(O-P) \\ (A+B)-(E+F) & (C+D)-(G+H) & (A-B)-(E-F) & (C-D)-(G-H) \\ (I+J)-(M+N) & (K+L)-(O+P) & (I-J)-(M-N) & (K-L)-(O-P) \end{bmatrix}$$

(c)

Figure 5: (a) The source image (b) Row operation in 2-D Haar DWT (c) Column operation in 2-D Haar DWT

A sample of 4x4 grey level images is shown in Fig 5(a). The addition and subtraction is applied on grey scale image for evaluating wavelet coefficient. The two dimensional discrete wavelet transform is accomplished by dual structured one dimensional discrete wavelet transform with both rows and columns. The row operation is conducted first in order to obtain the output as shown in Fig 5(b). Column operation is then used for transformation which finally gives the output of two dimensional Haar discrete wavelet transform as shown in Fig 5(c). A gray-scale image is converted to one mean constituent sub-band and three detail constituent sub-bands using two dimensional Haar DWT. Using Haar discrete wavelet transform on the image, diversified information about the text regions can be identified from the sub-bands details. For an example, LL subband identifies mean constituents, HL sub-bands identifies vertical boundaries, LH sub-bands identifies horizontal boundaries, and HH sub-bands identifies diagonal boundaries. The easy way to understand this is to observe the Fig 4 (a) which is basically a grey-scale image when subjected to Haar discrete wavelet transform gives the output as represented in Fig 5. The candidate text boundaries in the source image can seem from the detailed constituent's sub-bands (HL, LH, and HH).

B. K-Means Clustering

The k-means is basically a clustering algorithm which partition a data set into cluster according to some defined distance measure [24][25]. One of the significant tasks in machine learning is to comprehend images and extracting the valuable details. In this direction of analyzing data within the image, segmentation is the first phase to estimate quantity of the object present in an object. K-means clustering algorithm is an unsupervised clustering protocol [25] which categorizes the input data points into multiple types based on their inherent distance from each other. The protocol considers that the data features create a vector space and tries to locate normal clustering in them. The K-means function is given in (1). [*mu*, *mask*]=kmeans (ima, k) (1) where mu is the vector of class means, mask is the classification image mask, *ima*is the color image and *k* is the number of classes. The points are clustered around centroids in eq. (2) which are obtained by minimizing the objective [25]. Let *m* = max (*ima*)+1 , then mu = {(1:k) * m} / (k+1) (2) The maximum function shown above is the maximum value in the in *ima*matrix which represents the colored image in order to achieve the maximum value of the content colors where the color values are revealed as a unit value for all pixel. This stage is done to explicitly describe the maximum number of levels that can be used for estimating the histogram. The working principle of the k-means clustering algorithm in the proposed system is as discussed below: i. The histogram of intensities which should highlight estimates of pixels in that specific tone is estimated as shown below(3) where, n = total estimates of observations k = total estimates of tones. Σ==*kiimn*1

The quantity of the pixels is estimated by the m_i which has equivalent value. The graph created with the help of this is only the alternative way to represents histogram. ii. The centroid with k arbitrary intensities as in eq. (2) should be initialized. iii. The following steps are iterated until the cluster labels of the image do not alters anymore. iv. The points based on distance of their intensities from the centroid intensities are clustered. v. The new centroid for each of the clusters is evaluated.

C. Morphological Operation

The morphological operations like dilation and erosions are used for better approach of refining text region extraction. The non-text regions are removed using morphological operations. Various types of boundaries like vertical, horizontal, diagonal etc are clubbed together when they are segregated separately in unwanted non-text regions. But, it is also known that the identified region of text consists of all these boundary and region information can be the area where such types of boundaries will be amalgamated. The boundaries with text are normally short and are associated with one other in diversified directions. The proposed system has deployed both dilation and erosion for associating separated candidate text boundaries in every detail constituent sub-band of the binary image.

Figure 7: Implementation of Morphological operations on three binary regions

Finally, the morphological operations like dilation and erosion is designed exclusively to fit use-defined input of text based image with various types of characteristics.

III. PROPOSED METHOD

The proposed work is designed to accept the input as an image where the final effective output is obtained as extracted text using k-means clustering algorithm and mathematical morphological operations. For contrast in the results, discrete wavelet transform is applied for decomposing the image to sub-bands at various scales with diversified resolution. The text area is considered as special texture with unbalanced texture charecteristics. Various statistical features like mean, standard deviation, and energy is estimated when the image with text is subjected to discrete wavelet transformation algorithm. After the image is subjected to wavelet transform, classification based on region is applied for compacting the text area within the scope of image. A specific sliding window is designed which reads the high frequency sub bands by sliding steps. The application can be considered that the dimension of each sub-band is M×N after subjecting one-level wavelet transform, and we have, d_1= mod (M-W, l_1), d_2= mod(N-H, l_2)

If d1 and d2 are not equal to zero, than it fails to superimpose all the area of every sub-band when sliding window reads the high frequency sub-bands by the step l_1 x l_2. The work also rejects all the contents which do not belong to the region. The statistical charecteristics of every sub-band is estimated. The process achieves 12 features by evaluating the charecteristics of three high frequency subbands. Finally 12- dimension text feature vector is constructed.

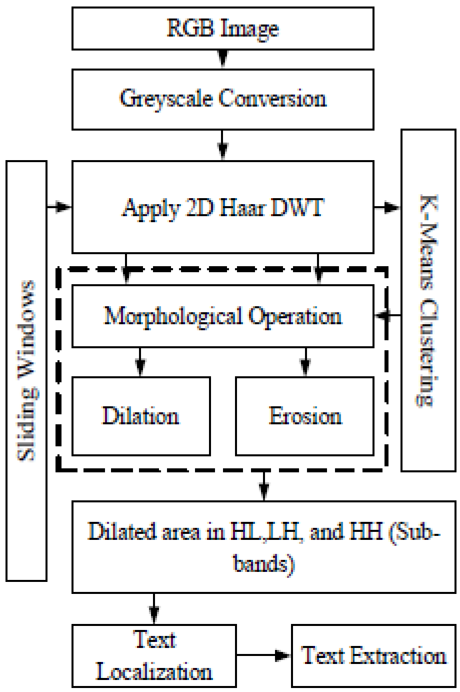

Figure 8: Overall Architecture of the proposed system

The second phase of the design uses k-means clustering protocol where clustering is deployed by analyzing the texture characteristic vector. The clustering factors selected are primary point of text, normal background, and complex background. Care should be taken to update the point of cluster in every processing of k- iterations. The image is segregated into three categories for textual area, simple and complex background area. Binarization technique is applied to the image depending on the results of classification and then mathematical morphological operations are deployed to take out the text details from the image. The effective algorithm implemented in the proposed system is as follows:

START
1. Input RGB image
2. If image is RGB
3.then covert to Gray scale
4. Create a function for performing DWT
5. Use Haar 2D DWT
6. Perform DWT
7. Initialize the coefficients, sub-bands
8. Create a function for sliding window
9. [W H] =size (window1)
10. mu = mean (mean (window1))
11. window2 = (window1-mu)
12. stanDev= sqrt (sum (sum (window2.^2))/(W*H))
13. E = sum (sum (window1.^2))
14. Estimate Size of subband
15. Create a function for K-Means Clustering
16. Calculate column number and row number
17. For zero padding
18. Apply zero Padding
19. Extract the features of sliding window
20. Rebuild the cluster id
21. Apply Mask Operation
22. Morphological operations on binary images
23. Detect boundary using Sobel
24. Morphologically open binary image (remove small objects)
STOP

One of the prime issues of implementing clustering algorithm is an inevitable computing error for which reason once the text area is extracted, the system cannot facilitate wholesome error free information about the complete text area. Therefore, the design implements morphological operations like erosion and dilation in order to measure and localize the all text sub-areas. Another issue is the non-text pixels which are also eliminated using erosion and dilation. The appropriate position of the text region is localized in the original image by merging the text pixel locus that is not extracted around the text region boundary. Finally the actual text information is extracted from the processed binarised image.

A. Text Extraction Pipeline

A text extraction system typically consists of five steps: (1) Localization, (2) Tracking, (3) Enhancement, (4) Binarization and (5) Recognition. The first step (Localization) aims to detect and accurately locate all the text lines in an image or a video frame. The second step (Tracking) helps to track the movement of the text lines over multiple frames, e.g., a text line moving from bottom to top in a movie credits scene. In the third step (Enhancement), the localized and tracked text lines are enhanced in terms of contrast and resolution to improve their readability. The fourth step (Binarization) converts the text lines into black and white images so that they can be used in the last step (Recognition), which recognizes the characters by using either an existing OCR engine or a custom-built OCR engine with its own feature extraction scheme.

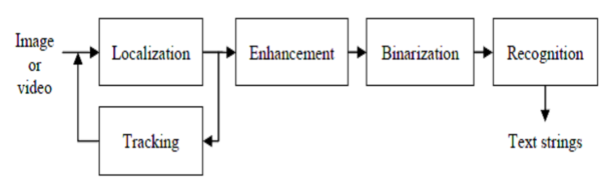

Figure 9: The typical steps of a text extraction system

Some text extraction systems may slightly change the order of the steps or omit certain steps. For example, Binarization is not needed if the Recognition step can work on grayscale or color images directly. As another example, because temporal information is not available in natural scene images, the Tracking step is omitted for these images. The next section discusses Localization, the first step in the pipeline.

B. Text Localization

The goal of text localization is to locate all the text lines in an input image or a video frame. A text line‗s position is usually represented by a rectangular bounding box .Some methods may provide additional information about a localized text line, e.g., a ―text mask‖, which indicates whether a particular pixel in the bounding box is a text pixel or a background pixel. Depending on the application, localization can also be performed at the word level, instead of at the text line level.

Different methods make use of different properties to localize the text lines. They can be classified into three main approaches: gradient-based, intensity/color-based and texture-based. As its name suggests, the first approach relies on the first two properties of text and often performs edge detection to identify regions in the input image with those properties. Similarly, the second approach analyzes regions in which the pixels have similar intensity values or colors (the third text property). Different from the previous two approaches, the last approach considers text as a special texture and applies techniques such as Discrete Cosine Transform and wavelet decomposition for feature extraction. For text/non-text classification, this approach typically employs machine learning techniques such as neural networks and Support Vector Machines (SVM).It is worth mentioning that unlike the first three properties, the last three properties of text are usually used at a later stage in a localization method (rather than as the main feature). For example, these properties can be used to remove false positives.

C. Gradient-based Localization

Gradient-based methods assume that in order for text to be readable, it needs to have enough contrast with the local background. Therefore, these methods look for regions with high intensity variation and/or dense edges. In addition, while most methods make use of ―unstructured‖ edges (e.g., in the form of edge energy or edge density), a few recent works focus on ―structured‖ edges such as strokes (parallel edges) and corners (intersected edges). .

Other than edge-related features, the property of high intensity variation in text regions has also been explored for text localization. (Kim & Kim 2009) made an interesting observation that due to color bleeding, there were often ―transient‖ pixels between text and background. These pixels were identified as groups of 3 consecutive pixels that followed an exponential increase/decrease in intensity values (depending on whether text was brighter/darker than the background). Region growing were performed to extend the transient pixels into candidate text regions. This method offers a new perspective into the problem of text localization and handles video graphics text well. However, it can only localize horizontal text and fails to pick up scene text, as shown in the sample results in the paper.

D. Texture-based Localization

To overcome the problem of complex background of gradient-based methods, the texture-based approach considers text as a special texture. These methods apply techniques such as Discrete Cosine Transform and wavelet decomposition for feature extraction. For text/non-text classification, they often employ machine learning techniques such as neural networks and SVM.

IV. EXPERIMENTAL RESULTS

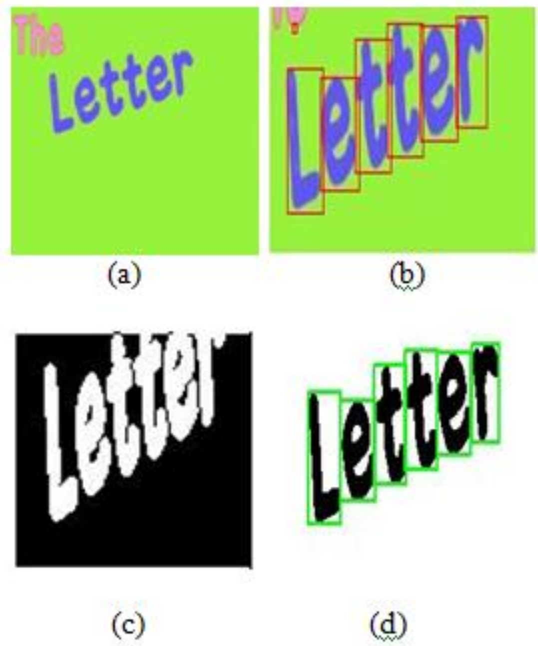

Figure 10: (a) Original Image (b) Segment character image (c) binarizaton output image (d) final segment output image

CONCLUSION AND FUTURE WORK

We have proposed a novel idea of combining Laplacian with wavelet high frequency sub-bands through fusion at multi-level to identify text candidates. We have explored Maximally Stable Extremely Regions along with stroke width distances for preserving fine details of text candidates. The proposed approach introduces mutual nearest neighbor clustering based on geometrical properties of text candidates to group text candidates of respective text lines into clusters. The symmetry driven growing process is proposed to extract arbitrary text lines based on the distance between text candidates in each cluster. According to our knowledge, this is the first attempt to detect text in both video frames and natural scene images with good accuracies. However, according to the results, the accuracy is still lower than that in document analysis, where usually the value would be more than 80%. We are planning to improve the accuracy of the current approach by making use of temporal frames in video through tracking in our future work.

References

[1] Palumbo, P. W., Srihari, S. N., Soh, J., Sridhar, R. and Demjanenko, V., 1992. Postal address blocks location in real time. Computer 25(7), pp. 34–42.

[2] Arth, C., Limberger, F. and Bischof, H., 2007. Real-time license plate recognition on an embedded DSP-platform.IEEE International Conference on Computer Vision and Pattern Recognition (CVPR ’07) pp. 1–8.

[3] Wolf, C., Michel Jolion, J. and Chassaing, F., 2002. Text localization, enhancement and binarization in multimedia documents. In: In Proceedings of the International Conference on Pattern Recognition (ICPR) 2002, pp. 1037–1040.

[4] Kavallieratou, E., Balcan, D., Popa, M. and Fakotakis, N., 2001. Handwritten text localization in skewed documents. In: International Conference on Image Processing, pp. I: 1102–1105.

[5] Wahl, F., Wong, K. and Casey, R., 1982. Block segmentation and text extraction in mixed text/image documents. Computer Graphics and Image Processing 20(4), pp. 375–390.

[6] KeechulJunga, Kwang in Kimb, Anil K. Jainc, Text information extraction in images andvideo: a survey, Pattern Recognition 37 (2004) 977 – 997

[7] H.J. Zhang, Y. Gong, S.W. Smoliar, S.Y. Tan, Automatic parsing of news video, Proceedings of IEEE Conference on Multimedia Computing and Systems, Boston, 1994, pp. 45–54.

[8] A.W.M. Smeulders, S. Santini, A. Gupta, R. Jain, Content-basedimage retrieval at the endof the early years, IEEE Trans. Pattern Anal. Mach. Intell. 22 (12) (2000) 1349–1380.

[9] M.A. Smith, T. Kanade, Video skimming for quick browsing basedon audio andimage characterization, Technical Report CMU-CS-95-186, Carnegie Mellon University, July 1995.

[10] M.H. Yang, D.J. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 34–58.

[11] Syed SaqibBukhari, Coupled Snakelet Model for Curled Textline Segmentation of Camera-Captured Document Images, Proceedings of the 2009 10th International Conference on Document Analysis and Recognition, 2009

[12] Samuel Dambreville, Tracking deformable objects with unscented Kalman filtering and geometric active contours, American Control Conference, IEEE, 2006

[13] Xu. C, Prince. J, "Snakes, shapes, and gradient vector flow" , IEEE Transaction on Images Processing, Vo!. 7, 1998, pp. 359-369

[14] Wumo Pan, T.D. Bui, C.Y. Suen, Text detection from natural scene images using topographic maps and sparse representations, IEEE, 2009

[15] J. Fabrizio, M. Cord, B. Marcotegui. Text Extraction from Street Level Images, CMRT09 - CityModels, Roads and Traffic 2009. Paris, France.

[16] Baba, Y.[Yoichiro], Hirose, A.[Akira], Spectral Fluctuation Method: A Texture-Based Method to Extract Text Regions in General Scene Images, IEICE(E92-D), No. 9, September 2009, pp. 1702-1715

[17] G. Aghajari, J. Shanbehzadeh, and A. Sarrafzadeh, A Text Localization Algorithm in Color Image via New Projection Profile, Proceedings of The International Multi conference of Engineers and Computer Scientist, Vol-2, 2010

[18] HrvojeDujmić, MatkoŠarić, JoškoRadić, Scene text extraction using modified cylindrical distance, Proceeding NNECFSIC'12 Proceedings of the 12th WSEAS international conference on Neural networks, fuzzy systems, evolutionary computing & automation, World Scientific and Engineering Academy and Society (WSEAS), 2011.

[19] Jayant Kumar; Rohit Prasad; Huiagu Cao; WaelAbd-Almageed; David Doermann; PremkumarNatarajan, Shape codebook based handwritten and machine printed text zone extraction, Proceedings Vol. 7874, Document Recognition and Retrieval XVIII, 2011

[20] SumitVashishta, Yogendra Kumar Jain, Efficient Retrieval of Text for Biomedical Domain using Data Mining Algorithm, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 4, 2011

[21] R. Pradip Kumar, P, Nagabhushan, Multiresolution Knowledge Mining using wavelet transform, Engineering Letter, 2007

[22] Abhayaratne, G.C.K., Discrete wavelet transforms that have an adaptive low pass filter, Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on Issue Date: 1-4 July, Vol 2, 2003

[23] Patrick J. Van Fleet, Discrete Haar Wavelet Transforms, PREP - Wavelet Workshop, 2006

[24] S K Gupta, K SambasivaRao and VasudhaBhatnagar, K-means Clustering Algorithm for Categorical Attributes, DataWarehousing and Knowledge Discovery Lecture Notes in Computer Science, 1999

[25] HassanaGremaKaganami, ZouBeiji, M Sami Soliman, Optimal Color Image Enhancement Using Wavelet and K-means Clustering, International Journal of Digital Content Technology and its Applications. Volume 5, Number 1, January 2011

[26] Y.M.Y. Hasan, L.J. Karam, Morphological text extraction from images, IEEE Trans. Image Process. 9 (11) (2000) 1978–1983.