ISSN 2278-3083

Volume 7, No.1, January – February 2018

International Journal of Science and Applied Information Technology

Available Online at http://www.warse.org/ijsait/static/pdf/file/ijsait01712018.pdf

A Novel Adaptive method to extract Text Information from Images for Information Retrieval

Dr V S Giridhar Akula [1], **Smd Mujeeb**[2]

Professor and Principal, Bandari Srinivas Inst of Technology, Hyderabad, Telangana, India.akulagiri2002@yahoo.com

[2]Associate Professor MallaReddy Engineering College, Hyderabad, Telangana, .mujeeb.smd@gmail.com

ABSTRACT

Many approaches are identified in extracting the text information represented in the natural images. The process of extracting the text from images applies with detection, tracking and recognition procedures. The extracted text can be used to retrieve the original images. Due to the differences in size, orientation and alignment of the text the process of extracting text is a challenging one. This paper focuses on text extraction procedures and thereby to retrieve the text by indexing the extracted text.

Key words : Text extraction, frames, page decomposition, pixels, polychorome, Caption Text, Gabor Filters, SVM.

1. INTRODUCTION

With an increasing demand for Internet and Intranet applications, there is a possibility of developing large amounts of data and extracting text from images and videos is a challenging task. Text data present in images contain important information which will be used for indexing, annotation and structuring of images. Many approaches have been proposed on extracting the data from natural images. Text Extraction procedure agrees an input as a still image or a sequence of frames. The frames can be in gray scale or color, compressed or un-compressed, and also the text within the frames. Text in frames will exhibit many variations according to their properties such as Geometry (size, alignment, inter-character distance), Color(monochrome, polychrome), Motion(static, linear moment), Edge(text boundaries, strong edges), Compression, etc.

Text extraction is generally divided into two components, detection and recognition. Scene text detection is to find out the regions containing text from camera captured images/videos. Text layout analysis based on gradient and color analysis is performed to extract candidates of text strings from cluttered background in natural v scene. Then text structural analysis is performed to design effective text structural features for distinguishing text from non-text outliers among the candidates of text strings. Scene text recognition is to transform image-based text in detected regions into readable text codes

Text present in images can have many differences with respect to the following properties:

1. **Geometry of the Image**:

   Text Size: Text sizes will vary and so we presume depending on the application domain.

   Text Alignment: The characters in the caption text appear in clusters and usually lie horizontally, although sometimes they can appear as non-planar texts as a result of special effects. This does not apply to scene text, which can have various perspective distortions. Scene text can be aligned in any direction and can have geometric distortions

   Inter-character distance: characters in a text line have a uniform distance between them.

2. **Color**: The characters in a text line tend to have the same or similar colors. This property makes it possible to use a connected component-based approach for text detection. Most of the research reported till date has concentrated on finding 'text strings of a single color (monochrome)'. However, video images and other complex color documents can contain 'text strings with more than two colors (polychrome)' for effective visualization, i.e., different colors within one word.

3. **Motion**: Characters usually exist in consecutive frames in a video with or without movement. This nomenclature is used in text tracking and enhancement. Caption text usually moves in a uniform way ie either horizontally or vertically. Scene text can have arbitrary motion due to camera or object movement.

4. **Edge:** in general, scene text are designed to be easily read, thereby resulting in strong edges at the boundaries of text and background.

5. **Compression**: Most of the digital images are recorded, transferred, and processed in a compressed format. Thus, a faster TIE system can be achieved if one can extract text without decompression.

2. LITERATURE REVIEW

Tang et al. explained a survey of page layout analysis, while Jain and Yu provided a brief survey of page decomposition techniques. Smith and Kanade identified the text information based on scene-change which allows the distinction between two consecutive. Banerjee et al. adopted the consistency of text characters in different sections to restore document images from severe degradation based on the model of Markov Random Field. Lu et al. suggested a word shape coding scheme through three topological features of characters for text recognition in document image.

Few researchers suggested the extraction of text from cluttered backgrounds. Kasar et al. assigned a bounding box to the boundary of each candidate character in the edge image and then detected text characters based on the boundary model. Sobottka et al. integrated a top-bottom analysis based on color variations in each row and column with a bottom-top analysis based on region growing by color similarity.

Shim et al. proposed the homogeneity of intensity of text regions in images. CC based methodology was adopted by Jain and Yu, which incorporates bit dropping, color clump, ambiguous image decomposition, and foreground image generation. Once the input frame is decomposed into multiple foreground frames, every foreground image goes through identical text localization stage.

Smith and Kanade focused on scene-change based on the difference between two consecutive frames and then used this scene-change information for text detection. They achieved an accuracy of 90% in scene-change detection. Lim et al., made a simple assumption that text usually has a higher intensity than the background. They counted the number of pixels that are lighter than a predefined threshold value and exhibited a significant color difference relative to their neighborhood.

Stroke Width Transform was proposed in 2010 , in which a local operator associated with stroke width is designed to model the specific structure of text character and extract text character components from non-text background outliers. In this procedure if a pixel is labeled more than once, the minimum length value will be assigned to it. Then pixels labeled by similar stroke width values are grouped as candidate connected components of text characters.

While employing the extraction of text string fragment, many methods were proposed. Akoum et al. proposed gradient-based analysis to localize and recognize car license plates. Ma et al. & Zhang et al. developed multi-scale edge features detect scene text, and edge-based features are also adopted in our framework to model text-specific structure. In other representations , a group of geometrical constraints was defined to link neighboring candidate character components that probably belonged to the same text string.

Many previous works have adopted machine learning tools and structural feature designs for modeling of text structures from fragments. Hanif et al. used mean difference, standard deviation, and HOG features of text characters to generate text detector under a Complexity Adaboost model. In the other related work, the responses of globally matched wavelet filters from text 44 regions are used as features to train text classifier based on Support Vector Machines (SVM) model and Fisher model. Gabor filter was used to segment text from documents. Pan et al. applied steerable Gabor filters to extract rotation-invariant features of multiple scripts. Shi et al. used gradient based curvatures to perform structural analysis of handwritten digits under a Bayes discriminant model.

Hase et al. proposed a CC-based method for color documents. They assume that every character is printed in a single color. Scene Text Character recognition, which generally includes feature representation to model character structure and multi-class classification to predict label and score of character class. J. Weinman, E. Learned-Miller, and A. Hanson, used Gabor filter responses on synthetic STC were employed to extract features of character appearance. Then the results of STC prediction are combined with language, similarity and lexicon model to perform word-level recognition. D. Smith, J. Feild, and E. Learned-Miller, SIFT descriptors were adopted to build a similarity expert to compute the character similarity, based on which integer program was applied for word recognition.

Hasan and Karam [42] presented a morphological approach for text extraction. The RGB components of a color input image are combined to give an intensity image Y as follows:

$$Y = 0.299\,R + 0.587\,G + 0.114\,B\ ,$$

where R, G, and B are the red, green, and blue components, respectively.

P. Viola and M. Jones used an efficient machine learning algorithm known as Cascade-Adaboost classifier in real-time face detection . The training process is divided into several stages. In each stage, a stage-specific Adaboost classifier is learned from a training set, which consists of all positives and the negatives incorrectly classified by previous Adaboost classifiers at this stage.

Many works have witnessed good results on Optical Character Recognition (OCR) systems. Tran et al. addressed ridge points in different scales to describe text skeletons at the level of higher resolution and text orientations at the level of low resolution; Liu et al. developed a stroke filter to extract the stroke-like structures. Wolf et al. modified Otsu's method to binarize text regions from background

3. EXTRACTION AND ENHANCEMENT OF THE TEXT

Temporal changes in the frame sequence is required to enhance the system performance. The text tracking step can serve to verify the text localization results. Along with that, if text tracking could be performed in lesser time than text detection and localization, this would speedup the overall system.

For the cases where text is blocked in different frames, text tracking can help recover the original image. Lienhart described a block-matching algorithm, which is an international standard for video compression such as H.261 and MPEG, and used temporal text motion information to refine extracted text regions. The mean absolute difference is used as the matching criterion. All localized blocks are checked as to whether its fill factor is above a given threshold value. For every block that satisfies the required fill factor, a block-matching algorithm is performed. If a block has an equivalent in a subsequent frame and the gray scale difference between the blocks is less than a threshold value, the block is considered as a text component.

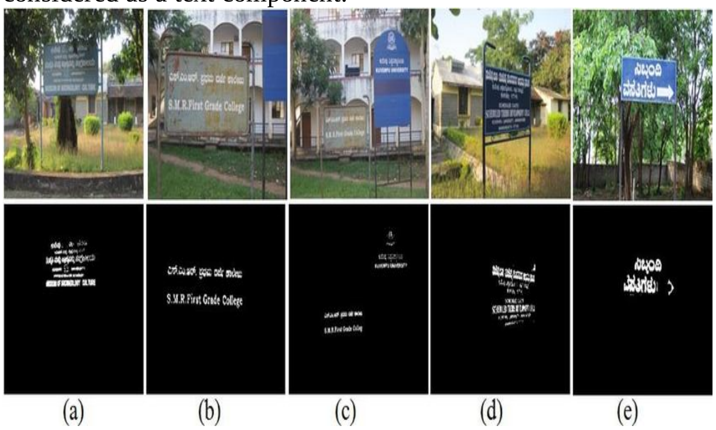

Fig: 1 sample images on Text extraction from the images (courtesy: researchgate.net)

Optical Character Recognition systems can produce an extremely high recognition rate for machine-printed documents on a simple background. Liang et al. implemented texture flow analysis to perform geometric rectification of the planar and curved documents. Burns et al. adopted performed topic-based partition of document image to distinguish text, white spaces and figures. Banerjee et al. adopted the consistency of text characters in different sections to restore document images from severe degradation based on the model of Markov Random Field. Lu et al. suggested a word shape coding scheme through three topological features of characters for text recognition in document image But, it is not easy to use commercial OCR software for identifying text extracted from images or video frames. Novel OCR systems need to be developed to handle large amount of noise and distortion in text extraction applications. Sawaki et al. proposed a method for adaptively acquiring templates of degraded characters in scene images involving the automatic creation of 'context-based image templates' from text line images. Zhou et al. use their own OCR algorithm based on a surface fitting classifier and an n-tuple classifier.

Camera-oriented scene images generally have complex background filled with non-text objects in multiple shapes and colors. In these images, text strokes, characters, and strings keep conspicuous by consistent colors. Thus many color-based clustering methods of text localization and text segmentation are designed in . However, these clustering methods ignore the color differences among neighboring pixels around the object boundaries. Compared with absolute color value that describes only a planar region, color difference value covers the color information of two neighboring planar regions.

4. CONCLUSION

We have presented the comprehensive literature review of text extraction in images as well as text based image and video retrieval. More number of algorithms have been presented but didn't find any techniques which provide satisfactory performance. The different information Sources (e.g., color, texture, motion, shape, geometry, etc). are used for text. SWT, OCR and machine learning algorithms were also used by many researchers to extract text from scene images.

.

REFERENCES

1. Y. Cui and Q. Huang, **Character Extraction of License Plates from Video**, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 502 –507.

2. H. Li and D. Doermann, **A Video Text Detection System based on Automated Training**, *Proc. of IEEE International Conference on Pattern Recognition*, 2000, pp. 223-226.

3. A. K. Jain, and Y. Zhong**, Page Segmentation using Texture Analysis**, Pattern Recognition, 29 (5) (1996) 743-770.

4. B. T. Chun, Y. Bae, and T. Y. Kim **Automatic Text Extraction in Digital Videos using FFT and Neural Network**, *Proc. of IEEE International Fuzzy Systems Conference, 1999*, Vol. 2, pp. 1112 –1115.

5. Y. Y. Tang, S. W. Lee, and C. Y. Suen, **Automatic Document Processing: A Survey**, Pattern Recognition, 29 (12) (1996) 1931-1952.

6. B. Yu, A. K. Jain, and M. Mohiuddin, **Address Block Location on Complex Mail Pieces**, *Proc. of International Conference on Document Analysis and Recognition*, 1997, pp. 897-901.

7. J. C. Shim, C. Dorai, and R. Bolle, **Automatic Text Extraction from Video for Content-based Annotation and Retrieval**, *Proc. of International Conference on Pattern Recognition*, Vol. 1, 1998, pp. 618-620.

8. K. Jung, K. I. Kim, and J. Han, **Text Extraction in Real Scene Images on Planar Planes**, *Proc. of International Conference on Pattern Recognition,* 2002, Vol. 3, pp. 469-472.

9. S. Antani, **Reliable Extraction of Text From Video**, PhD thesis, Pennsylvania State University, August 2001

10. M.A. Smith and T. Kanade, **Video Skimming for Quick Browsing Based on Audio and Image Characterization**, Technical Report CMU-CS-95-186, Carnegie Mellon University, July 1995

11. U. Gargi, D. Crandall, S. Antani, T. Gandhi, R. Keener, and R. Kasturi, **A System for Automatic Text Detection in Video,** *Proc. of International Conference on Document Analysis and Recognition*, 1999, pp. 29 –32.

12. Yu Zhong, Hongjiang Zhang, and Anil K. Jain, **Automatic Caption Localization in Compressed Video**, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, (4) (2000) 385-392.

13. J. Ohya, A. Shio, and S. Akamatsu, **Recognizing Characters in Scene Images**, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16 (2) (1994) 214-224.
14. M. Pilu, **On Using Raw MPEG Motion Vectors to Determine Global Camera Motion**, *Proc. of SPIE*, 1998, Vol. 3309, pp. 448-459. 64. H. Li, O. Kia, and D. Doermann, Text Enhancement in Digital Vi
15. E. Y. Kim, K. Jung, K. Y. Jeong, and H. J. Kim, **Automatic Text Region Extraction Using Clusterbased Templates**, *Proc. of International Conference on Advances in Pattern Recognition and Digital Techniques*, 2000, pp. 418-421 *Advances in Vision Computing: An International Journal (AVC*) Vol.1, No.1, March 2014.
16. J. Banerjee, M. Namboodiri, and C. Jawahar, **Contextual restoration of severely degraded document images**, *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 517-524.
17. T. Burns and J. Corso, **Robust unsupervised segmentation of degraded document images with topic models** , *IEEE Conference on Computer Vision and Pattern Recognition,* 2009, pp. 1287-1294.
18. A. Akoum, B. Daya, and P. Chauvet, **A new algorithmic approach for detection and identification of vehicle plate numbers**, *Journal of Software Engineering and Applications*, pp. pp. 99-108, 2010.
19. B. Epshtein, E. Ofek, and Y. Wexler, **Detecting text in Natural scene with stroke width transform**,*IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
20.T. Kasar, J. Kumar, and A. Ramakrishnan, **Font and background color independent text binarization**, Camera-based Documentation Analysis and Recognition, 2007, pp. pp. 3-9.
21. S. Hanif and L. Prevost, **Text detection and localization in complex scene images using constrained Adaboost algorithm**, *International Conference on Document Analysis and Recognition,* 2009, pp. 1-5.
22. S. Lu and C. Tan, **Retrieval of machine-printed Latin documents through word shape coding**, *Pattern Recognition*, vol. 41, no. 5, pp. 1799-1809, 2008.