ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 5, Issue 5, May 2016

Comparative Study and Performance Evaluation of Various Text Extraction Techniques in Image

Soumen Bhowmik, Aloke Kumar Mondal

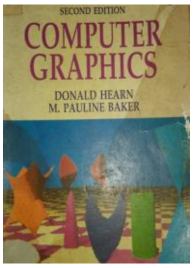
***Abstract*— Most of the time information's are written on an image. Sometime this information's are so useful to extract. Text extraction is one of the key tasks in document image analysis. Automatic text extraction without characters recognition capabilities is to extract regions just contains text. The text extraction process includes detection, localization, segmentation and enhancement of the text from the given input image. In this paper we present a comparative study and performance evaluation of various text extraction techniques.**

Index Terms— **Text Detection, Text Localization, Text Tracking, Text Enhancement, Text Extraction, OCR**

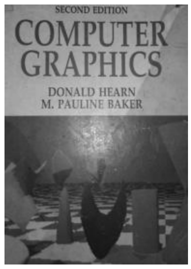
I. INTRODUCTION

Text Extraction from image is concerned with extracting the relevant text data from a collection of images. Text extraction from an image is a challenging problem because of image contains text of different size, style, orientation, alignment, low contrast, noise and has complex background structure. Content-based image indexing refers to the process of attaching labels to images based on their content. Image content can be divided into two main categories: perceptual content and semantic content [1]. Perceptual content includes attributes such as color, intensity, shape, texture, and their temporal changes, whereas semantic content means objects, events, and their relations. Since the text data present in image or video in different variations, the problem of extracting the text region from images becomes a challenging one. Among them, text within an image is of particular interest as (i) it is used to describe the contents of an image; (ii) it can be easy to extract as compared to other semantic contents, and (iii) it enables applications such as keyword-based image search text-based image indexing etc.

Text in Images: A large number of approaches to text information extraction (TIE) from images have been proposed for specific applications and including page segmentation, address block location, license plate location, and content-based image/video indexing. In spite of such extensive studies, it is still not easy to design a general-purpose TIE system. 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. Some Example of Text in Images are given below.



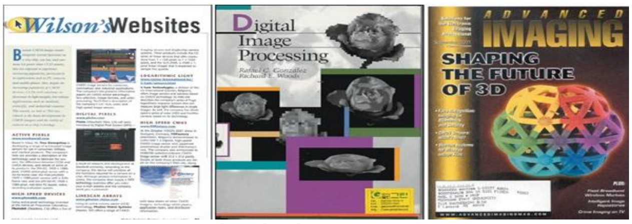
Normal Image



Gray Scale Image

Scene Text Image


Two column page from journal



Multi-color document image

II. PROPERTIES OF TEXT IN IMAGES

Due to the variations in geometrical properties of text and background complexities the problem of automatic text extraction extremely complicated and difficult. Text in images can exhibit many properties with respect to the following properties:

A. Size: Although the text size can vary a lot, assumptions can be made depending on the application domain.

B. Alignment: The caption texts 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 has various perspective distortions. Scene text can be aligned in any direction and can have

***Soumen Bhowmik**, Assistant Professor, Department of CSE, Bengal Institute of Technology and Management, Santiniketan, Phone/ Mobile No.09832157090, India.*

***Aloke Kumar Mondal**, M. Tech Student, Department of CSE, Bengal Institute of Technology and Management, Santiniketan, India, Phone/ Mobile No.09476134098.*

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

C. Edge: Most caption and scene text are designed to be easily read, thereby resulting in strong edges at the boundaries of text and background.

D. Color: In a simple image the characters in a text usually have the same or similar colors. This property makes it possible to use a connected component-based approach for text detection. However, video images and other complex color documents usually contain text strings with more than two colors.

E. Motion: The same characters usually exist in consecutive frames in a video with or without movement. This property is used in text tracking and enhancement. Caption text usually moves in a uniform way: horizontally or vertically. While, scene text can have arbitrary motion due to camera or object movement.

F. Compression: The digital images are usually processed in a compressed format. If one can extract text without decompression then it is possible to develop a faster Text Information Extraction (TIE) system.

III. TEXT INFORMAION EXTRACTION

A Text Information Extraction system (TIE) receives an input image and output the relevant text data. As stated in the images can be in gray scale or color, compressed or uncompressed. The Text Information Extraction (TIE) as shown in Fig.1includes the following stages -

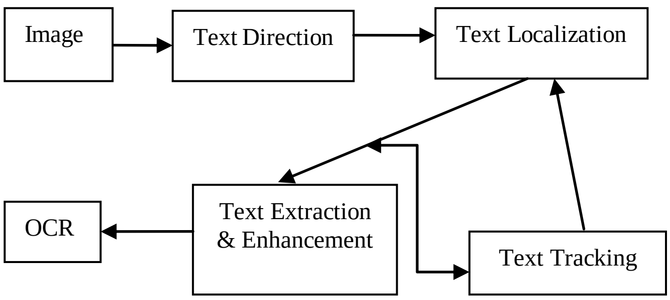


Fig.1: Text Information Extraction (TIE)

A. Text detection: Text detection refers to the process of determining whether or not the presence of text in a given image/frame.

B. Text localization: Text localization refers to the determination of the location of text in the image/frame and generating bounding boxes around the text.

C. Text tracking: Text tracking is used to reduce the processing time for text localization.

D. Text extraction: Text extraction is the stage where the text needs to be segmented from the background to facilitate its recognition.

E. Text Enhancement: Text Enhancement is required because when the text region is extracted from the background it usually has low resolution and is more likely to suffer from noise. Thereafter, the extracted text images can be transformed into plain text using OCR technology.

In spite of extensive studies, it is not easy to design a general-purpose Text Information Extraction (TIE) system. This is because there are many possible sources of variation when text is being extracted from the complex background, from low contrast or complex images or from the images having variations in geometric properties of text. These variations make the problem of automatic Text Information Extraction (TIE) extremely difficult.

IV. DIFFERENT TEXT EXTRACTION TECHNIQUES

Text extraction is one of the required stages prior to character recognition. The aim of text extraction is to separate each character so that it can be fed into the recognition stage. This paper discusses about different text extraction techniques such as region-based, edge-based, texture-based and morphological-based techniques.

A. Region-Based Techniques: Region-based methods use the properties of the color or gray scale in a text region or their differences with the corresponding properties of the background. This method uses a bottom-up approach by grouping small components into successively larger components until all regions are identified in the image. A geometrical analysis is needed to merge the text components using the spatial arrangement of the components so as to filter out non-text components and mark the boundaries of the text regions.

Debapritam Sarkar [12][2][3][4] described how text can be extracted from handwritten text using bottom-up approach of Line Segmentation. The process of Line segmentation is done in which the consecutive lines are extracted or separated from each other from a text. Line segmentation is mainly of two types (1) Line Segmentation of Typed text, (2) Line Segmentation of Handwritten text. The line segmentation of unconstrained hand written text is very much difficult because of interline distance variability and the base line skew variability. In Line segmentation of handwritten text, the picture is first divided into small squares with height and width 10 pixels each. If 50% of the square box is filled up with black pixels then the total square is filled with black pixels. If 3 or more squares around a white square are black then the white block is also made black to have better smoothing. In this way graphically smooth image is found. Thus graphically smooth image is computed with the height of each of components. Next a rectangular template is created with basis of a specified width and height as maximum portable height. The template scanned from the top left corner of the graphically smooth image down below by one pixel till the bottom of the image is reached. If the number of black pixel obtain for each position for a image is greater than the number of the previous position of the template the previous information is discarded. To get individual lines the smooth blocks are joined Depending on the height and the position information. Next the lines are extracted with the help of upper and lower boundaries. Then lines are placed one after another in a link list, i.e. the nodes of the link list are the lines.

Karim, Bunke and Kronenberg [11][2][4][3] describe a method for Identification of Text on Colored Book and Journal Covers. From here we can learn how to minimize color variations by applying clustering methods in preprocessing step. They proposed two methods for extracting text hypotheses: a) Top down analysis: the image is split vertical and horizontal directions alternatively in this phase. From this top down approach always we get rectangular shaped regions only. Regions containing text include at least two colors. Depending on this information we reject the regions having no text. b) Bottom up analysis: Region growing method is used to detect homogeneous regions. Beginning with a starting pixel, pixels are merged if they are belonging to the identical cluster. We know that characters of printed text generally do not touch each other so several regions are detected for a text region. The results of both methods are combined to distinguish between text and non text regions. Text elements are binarized using automatically extracted information about text color. The binarized text regions used as input for a conventional OCR module. The proposed method is not restricted to cover pages, but can be applied to the extraction of text from other types of color images as well.

B. Edge-Based Techniques: an edge in the images is the unique features of text regardless of color/intensity, orientations, etc. Edges are considered as the distinguishing characteristics of text embedded in images, as it is the main features for detecting text. Edge-based text extraction algorithm is a general-purpose method, for effectively localize and extract the text from both indoor/outdoor images.

Yingzi Du [13][2] propose an edge-based technique consists of four modules: multistage pulse code modulation(MPCM) module, text region detection (TRD) module, text box finding (TBF) module and optical character recognition (OCR) module. In the first module we get coded image as MPCM convert a color video image into a gray-scale image and further produce a binary thresholded image, which locates potential text regions. In coded image each pixel encoded by a priority code ranging from 0 to 7 in accordance with its priority and further produces a binary threshold image. It was discovered that such a priority code provides a good measure for locating potential text regions. The TRD module uses spatial filters to remove noisy regions and it also eliminate regions that are unlikely to contain text. Five filtering steps are included in this module: thresholding, elimination of isolated blocks, elimination of long vertical blocks, elimination of diagonally connected blocks and elimination of weakly connected vertical blocks. The design of these spatial filters takes advantage of how text appears in images. Since a text region that contains characters almost always appears as a box. TBF module is to rectangularize the text regions detected by the TRD module and produce text boxes. By doing so, several text regions are merged to form a single text box. The final the OCR module is used to process each of the text boxes and to eliminate those boxes that produce no OCR results. From OCR module we can make simple binary decision to determine whether a text box contains text or not.

Xiaoqing Liu [14][2][3][4]proposes a multi-scale edge-based text extraction algorithm which can quickly and effectively localize and extract text from both documents and images. The proposed method consists of three stages: a) candidate text region detection: In this stage build a feature map by using three important properties of edges: i) edges strength ii) density iii) variance of orientations. A feature map is binary image where pixel intensity gives possibility of text. b) text region localization: In this stage morphological operator is used. There are two constraints utilized to find non text regions, first for finding very small isolated blocks and second for filter out the block whose width is very small than that's corresponding height .c) character extraction: Here existing OCR engines were used for character extraction. This can only deal with printed characters against clean backgrounds and cannot handle characters embedded in shaded, textured or complex backgrounds.

C. Texture-Based Techniques: Texture-based methods use the observation that texts in images have distinct textural properties that distinguish them from the background, to decide whether a pixel or block of pixels belong to text or not. Text feature extraction lies essentially on image pre-processing techniques, which is usually performed by linearly transforming or filtering the textured image followed by some energy measure or non-linear operator. The techniques based on Gabor filters, Wavelet, FFT, spatial variance, etc. can be used to detect the textural properties of a text region in an image.

Kwang [15][3][4] describes texture-based method for detecting texts in images. A support vector machine (SVM) is used to analyze the textural properties of texts. No external texture feature extraction module is used. The intensities of the raw pixels that make up the textural pattern are fed directly to the SVM, which works well even in high-dimensional spaces. Next, text regions are identified by applying a continuously adaptive mean shift algorithm (CAMSHIFT) to the results of the texture analysis. The combination of CAMSHIFT and SVMs produces both robust and efficient text detection, as time-consuming texture analyses for less relevant pixels are restricted. Leaving only a small part of the input image to be texture-analyzed. The performance criterion was the classification accuracy of the SVMs for text and non text patterns rather than the overall text detection results. For this purpose, 100 training images were divided into two different classes of 70 training images and 30 validation images from which training patterns and validation patterns were collected respectively. The SVMs were then trained using the training patterns and tested using the validation patterns.

Bassem Bouaziz [2] describes texture-based method for detecting texts in images. Based on local application of Hough transform combined with use of transformation matrix this method has four steps: a) Sweeping the image: Let S a set of collinear pixels forming a line segment within an image. Then the extremities of a given segment can be identified by sweeping sequentially the image from top to bottom and from left to right. b) Detecting segments: When a line segment is detected, it is stored and removed from the image. Then sequential search continues until the whole image is swept.

ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 5, Issue 5, May 2016

When a line segment extremity is reached, sweeping can be done in all directions to find direction where most connected pixels exists. In order to improve performances and avoid call of trigonometric functions, two transformation matrixes can be computed in the initialization step. c) Storing information about segments: So each element can be represented as a pixel coordinates in image space by using maximal length of line segment that detected in a direction between 0 and 180°, and by the coordinates of line segment extremity identified when sweeping the image. The obtained matrix represents neighbourhood's information of a detected extremity concerning connected pixels. This consideration of neighbourhood will help to detect imperfect segment as the case of an edge image. d) Detecting regularity: The last step of algorithm consists of removing segment's pixels that are having length exceeding a threshold value, which represents the minimal length of segment that should detect. Regularity can be detected if distance between parallel segments is similar for a specified value.

D. Morphological-Based Techniques: Mathematical morphology is a topological and geometrical based approach for image analysis. It provides powerful tools for extracting geometrical structures and representing shapes in many applications. Morphological feature extraction techniques have been efficiently applied to character recognition and document analysis. It is used to extract important text contrast features from the processed images. The feature is invariant against various geometrical image changes like translation, rotation, and scaling. Even after the lighting condition or text color is changed, the feature still can be maintained. This method works robustly under different image alterations.

Jui-Chen Wu [17][4][3][2] describe a text line extraction algorithm for extracting text regions from cluttered images. The method defines a set of morphological operations for extracting important contrast regions. The main steps of this algorithm are: a) Feature extraction: A noval morphology based scheme for extracting the high contrast feature for locating all possible text lines. Text lines embedded in images often have quite high contrast to the background. In addition their widths and heights are usually uniform and horizontally aligned. The relative contrast between texts and their background is an important feature for line detection. b) Text candidate selection: a labeling technique is used to select all possible text lines from the analyzed image. For skewed text lines a moment based method is applied to find their orientations. A novel rule based selection scheme is proposed to select all potential segments. A novel merging algorithm is proposed for merging the segment together if they belong to same text line. c) Candidate verification: After candidate selection a verification process is carried out. The text verifications done on the basis of regularities of character size, the ratio between character width and height, and the period of characters.

Rama Mohan [18][3][4] The method describes that edges are most reliable features of text than its color or intensity, layout, orientation etc. The edge detection operation is done using the basic operators of mathematical morphology. The algorithm is used to find out the connected component. By considering the gray levels of the components their variance is found out for each connected component, when components are found then labeling is done. After selecting the components whose variation is less than threshold value the text can be extracted. This method consists of four steps: a) Edge extraction: this method contains three stages i) Apply non-linear filter to the given input image to remove noise. ii) The blurred image obtained from the filtered image is taken as the input and we find the morphological gradient of this image. iii) Threshold is used to obtain binary image from the grayscale gradient image. b) Text candidate region formation: From the threshold image the text candidate regions are obtained as follows. In text candidate region formation close operation is applied to connect all the edges. c) Labeling of text candidate regions: Apply labelling on the text candidate region as follows • To the above obtained text candidate region each candidate is uniquely labelled. • Re-labeled the text candidate regions by sequentially assigning unique values to the same component. d) Elimination of non text region: From the labelled image which contains text and non text regions eliminate the non text regions using variance operation.

V. COMPARISONS AND PERFORMANCE EVALUATION

The performance evaluation of information retrieval can be done using precision and recall rate. The precision rate measures the percentage of correctly detected text boxes with in each image as opposed to detected boxes, where as percentage of correctly detected text boxes that actually contain in text are measured by recall rate.

Precision rate = Number of correctly detected text boxes / Number of detected text boxes

Recall rate = Number of correctly detected text boxes / Number of text boxes

Performance evaluation and comparison of different text extraction techniques discussed in this paper are listed in below table.

Technique	Method	Properties	Recall rate (%)	Precision rate (%)
Region Based	Bottom-Up Approach of Line Segmentation	Handwritten Text	92	
	Clustering, Top-Down successive splitting , bottom-up region growing	Coloured book and journal covers	83.3	83.5
Edge Based	Automated system for text extraction	Scene text and superimposed text within video images	92	85
	Multiscale Strategy , Clustering	Complex printed document images and scene text	96.6	91.8
Texture Based	Multiscale texture based method using local energy analysis	Hybrid Chinese/English text detection in images and video frames	93.5	1.8
	Hough Transform technique combined with an extremity	Detect text in video images	94	2.48

	segments neighbourhood analysis			
Morphological Based	A measure of accumulated gradients and morphological post processing	Artificial text in images and video	93.5	85.4
	Novel set of morphological operations and an x-projection techniques	Cluttered images	96.3	99.4

VI. CONCLUSION

We have provided a comprehensive survey of text information extraction in images. Even though a large number of algorithms have been proposed in the literature, no single method can provide satisfactory performance in all the applications due to the large variations in character font, size, texture, colour, etc. There are several information sources for text information extraction in images (e.g. colour, texture, motion, shape, geometry, etc). It is advantageous to merge various information sources to enhance the performance of a text information extraction system. It is, however, not clear as to how to integrate the outputs of several approaches. There is a clear need for a public domain and representative test database for objective benchmarking. The lack of a public test set makes it difficult to compare the performances of competing algorithms, and creates difficulties when merging several approaches.

REFERENCES

[1] Keechul Junga, Kwang In Kimb, Anil K. Jainc "Text information extraction in images and video: a survey" The Journal of Pattern Rechonization Society, Pattern Recognition 37 (2004) 977 – 997

[2] Kumary R Soumya, Ancy.T.V, Athulya Chacko "TEXT EXTRACTION FROM IMAGES : A SURVEY" International Journal of Advances in Computer Science and Technology,Volume 3, No.2, February 2014

[3] A.J.Jadhav, Vaibhav Kolhe, Sagar Peshwe "Text Extraction from Images A Survey " International Journal of Advanced Research in Computer Science and Software Engineering,Volume 3, Issue 3, March 2013

[4] C.P. SUMATHI, T. SANTHANAM, N. PRIYA "TECHNIQUES AND CHALLENGES OF AUTOMATIC TEXT EXTRACTION IN COMPLEX IMAGES A SURVEY" Journal of Theoretical and Applied Information Technology, 31st January 2012. Vol. 35 No.2

[5] Priyanka Deelip Wagh, D. R. Patil "Text Detection, Extraction and Removal A Survey diferent view" International Journal of Computer Applications (0975 – 8887),National Conference on Emerging Trends in Computer Technology (NCETCT-2014)

[6] Paraag Agrawal, Rohit Varma "Text Extraction from Images" IJCSET |April 2012| Vol 2, Issue 4,1083-1087

[7] J. Fabrizio, M. Cord, B. Marcotegui "TEXT EXTRACTION FROM STREET LEVEL IMAGES" In: Stilla U, Rottensteiner F, Paparoditis N (Eds) CMRT09. IAPRS, Vol. XXXVIII, Part 3/W4 --- Paris, France, 3-4 September, 2009

[8] Narasimha Murthy K N, Dr. Y S Kumaraswamy "A Novel Method for Efficient Text Extraction from Real Time Images with Diversified Background using Haar Discrete Wavelet Transform and K-Means Clustering" IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, September 2011

[9] C.P. Sumathi, T. Santhanam, G.Gayathri Devi. "A SURVEY ON VARIOUS APPROACHES OF TEXT EXTRACTION IN IMAGES" International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.4, August 2012

[10] Amit Choksi, Nihar Desai, Ajay Chauhan, Vishal Revdiwala, Prof. Kaushal Patel "Text Extraction from Natural Scene Images using Prewitt Edge Detection Method" International Journal of Advanced Research in Computer Science and Software Engineering,Volume 3, Issue 12, December 2013

[11] Karim Sobottka, Horst Bunke and Heino Kronenberg " Identification of text on coloured Book and Hurnal Covers"

[12] Debapratim Sarkar, Raghunath Ghosh "A Bottom-Up Approach of Line Segmentation from Handwritten Text"

[13] Yingzi Du, Chein-I Chang "Automated system for text detection in individual video images" Journal of Electronic Imaging 12(3), 410–422 (July 2003)

[14] Xiaoqing Liu and Jagath Samarabandu "Multiscale edge-based Text Extraction from complex image"

[15] Kwang In kim, Keechul Jung and Hyung Kim "Texture-Based Approch for Text Detection in Image Using Support BVector Machines and Continuous Adaptive Mean Shift Algorithm" IEEE Transaction on pattern analysis and machine intelligence Vol 25 No 12 Dec 2003

[16] Baseem Bouaziz, Tarek Zlitni, Walid Mahdi AViTExt "Automatic Video Text Extraction A new Approachfor video content indexing Application"

[17] Jui-Chen Wu, Jun-Wei Hsieh, yung-Sheng Chen "Morphology-based text line extraction" Machine vision and Application (2008) DOI 10.1007s00138-007-0092-0

[18] G.RAMA MOHAN BABU,P.SRIMAIYEE, A.SRIKRISHNA "TEXT EXTRACTION FROM HETROGENOUS IMAGES USING MATHEMATICAL MORPHOLOGY" Journal of Theoretical and Applied Information Technology 2005 - 2010 JATIT. All rights reserved.

[19] Wonjun Kim and Changick kim "A New Approach for Overlay Text Detection and Extraction From Complex Video Scene" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 18, NO. 2, FEBRUARY 2009

[20] Prachi R. Dussawar1, Prof. Parul Bhanarkar Jha "A SURVEY ON COMPARISION AND PERFORMANCE ANALYSIS OF TEXT EXTRACTION TECHNIQUES" International Journal of Advance Research In Science And Engineering IJARSE, Vol. No.4, Issue No.01, January 2015

[21] Prajakta D. Sawle, Chetan J. Shelke "Survey on Text Extraction in Android Mobile Application using Character Descriptor" International Journal of Advance Research in 1Computer Science and Management Studies, Volume 2, Issue 11, November 2014

[22] Renjie Jiang, Feihu Qi, Li Xu, Guorong Wu " Detecting and Segmenting Text from Natural Scenes with 2-Stage classification"

[23] C.P. Sumathi1, T. Santhanam and G.Gayathri Devi "A SURVEY ON VARIOUS APPROACHES OF TEXT EXTRACTION IN IMAGES" International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.4, August 2012

[24] Amit B. Kore and D. R. Mehta "TEXT AND AUDIO TRANSLATION OF TEXT FROM SIGNBOARD IMAGES – REVIEW" International Journal of Advances in Engineering & Technology, May, 2014.©IJAET ISSN 22311963

[25] M. Swamy Das,B. Hima Bindhu, A. Govardhan "Evaluation of Text Detection and localization methods in Natural Images" International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, Volume 2, Issue 6, June 2012)

[26] Vani Kasireddy,Swathi.G, Santhosh Kumar "An Extraction of overlay text from Digital Videos" International Journal of Engineering Research and Applications (IJERA) ISSN 2248-9622 NATIONAL CONFERENCE on Developments, Advances & Trends in Engineering Scie

[27] Vikas K. Yeotikar, Manish T. Wanjari, Dr. Mahendra P. Dhore "Text Extraction from Document Images Using Gabor, Wavelet and Hough Technique A Novel Approach" International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN 234

Soumen Bhowmik completed his B. Tech from University of Kalyani and M. Tech from IIEST, Shibpur, West Bengal. He has published several research papers in different National and International Journals. His areas of interest are image steganography, real time video data processing, sensor network etc. He is the Life member of ISTE, IACSIT-Singapore, CSTA-ACM, ISOC-Switzerland etc.

ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 5, Issue 5, May 2016

Aloke Kumar Mondal is pursuing M. Tech in the CSE branch under WBUT. He has completed his MCA from WBUT. He has the interest in different programming language, image processing, data base management System, networking and traffic control in networking.