International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 11 6105 - 6109

A Novel Approach to remove Ink Bleed through Degraded Document Images

Er. Hemani
Student,CSE Department
BFCET,Bathinda
India.
hemanichanana@gmail.com

Er. Charandeep Singh Bedi
Astt.Professor,CSE Department
BFCET,Bathinda
India
er_charandeep@yahoo.co.in

Abstract— There are numerous recorded reports which manages the corruption of paper because of paper maturing, foundation variety because of clamor, uneven brightening or dim spots and characterization of loss of literary data in debased archives like light introduction which causes blurring of content or ink chipping, Degradation of the written work medium i.e. clouded or missing content brought about because of the vicinity of mould,parasites,dampness or weakness in the medium,When ink has leaked through posterior or a page to front reasons seep through interference,whenever we digitize our record it may present some clamor curios which may make debasement the printed information.There are numerous corrupted yet truly vital old original copies and reports dispersed crosswise over libraries and chronicles the world over. Because of entry of time ink of rear begins to meddle with the ink of front side which hampers the intelligibility of archives. Be that as it may, because of the significance of such archives it is essential to restore such records. In this paper, different calculations are used in pre handling steps like Bernsen calculation, Improved Bernsen calculation, Canny edge discovery method are used for initialising the outcomes. Subsequently post preparing steps are proposed toward the end so that the calculation finishes up with improved and productive results.

Keywords- *Historical Documents, Ink-Bleed, Binarization, Bernsen algorithm, Improved Bernsen Algorithm, Canny Edge Detection technique.*

******

I. INTRODUCTION

Image restoration is a dynamic examination zone and different specialists work to enhance the proficiency of the diverse calculations by growing more productive calculations which keeps on being a testing exploration topic[1].There are a few purposes behind debasement which may happen oftentimes i.e. appearance of non-uniform force and low complexity. Debased report image binarization comes about more dependability as it is performed by evacuating unneeded articles that show up in the record, foundation concealing, clamor evacuation and filling the conceivable breaks in the lines, gaps or holes on the frontal area and enhances the last result with more nature of strokes inside of the characters before its transformation to editable content. Perceiving the chronicled records manages paper debasement because of paper maturing. On the other hand various sorts of clamors and corruption which incorporates variety in foundation, dull spots brightening which causes unevenness[22].

II. METHODOLGY

Initially of all,the archive is changed over into filtered image with the assistance of image scanner. The examined image can be in any arrangement for eg .jpg , .bmp , .tif, .png,etc. At that point comes the Binerization stage that intend's to segment a image's pixels into forefront and foundation bunch . It is center method in different applications, the edge extraction, and item extraction, preparing and image division shape, particularly. Binarization is a procedure of changing over image into twofold frame that implies as 0s and 1s.[23]

Further we utilize Bernsen's calculation which figures the neighborhood limit quality taking into account the mean estimation of the base and greatest intensities of pixels inside of a window. On the off chance that the window is focused at the pixel (x,y) the edge for I(x,y) is characterized by:

$$T(x,y) = \frac{(Zmax+Zmin)}{2}$$

where Zmax and Zmin are the most extreme and least force of the window. The difference is characterized as

$$C(x,y)= Z\ max – Z\ min.$$

If the differentiation is less that a particular quality k the pixels inside of the window may be set to foundation or to closer view as indicated by the class that most suitably depicts the window.

Here odd window size is utilized on the grounds that the calculation results with more viable images. Henceforth we take window size= 3*3 and the image ought to be two dimensional, i.e. dark scale image generally blunder will happen in light of the fact that the image would be three dimensional which is a RGB image.Further the image is changed over into twofold image so that the yield of the first image can store its outcome. Contrast edge is set to 15. Compute mean and nearby differentiation in order to intialize the output.

```
mean = averagefilter(image, window);
local_contrast = maxfilt – minfilt;
```

After this we add some more progressed to ventures to Bernsen calculation i.e, Calculating Mean worth utilizing normal channel furthermore utilizing exponential cases too for the outcomes to figure the nearby difference where r is for number of lines and c is for number of segments.

$$local\_contrast(i,j)=\frac{localcontrast(i,j)}{localcontrast2(i,j)}$$

Finally instating the yield and figure the cover i.e., if the complexity is less then edge the yield is 1 generally 0. At that point comes execution of watchful edge discovery calculation so we can without much of a stretch make out with the edges of the content written in the record.

Finally another strategy is presented utilizing Bernsen calculation, three post handling steps are added to the calculation i.e, Firstly ascertaining mean and mean square utilizing normal channel

```
mean = averagefilter(image, window, padding);
meanSquare = averagefilter(image.^2, window, padding);
```

Besides, standard deviation is ascertained which is utilized to evaluate the measure of variety or scattering of an arrangement of information qualities. A standard deviation near 0 demonstrates that the information focuses have a tendency to be near the mean (additionally called the normal worth) of the set, while an exclusive expectation deviation shows that the information focuses are spread out over a more extensive scope of qualities. Finally applying otsu technique which naturally perform's image thresholding or, the lessening of a graylevel image to a double image. The calculation expect that the image contains two classes of pixels that ascertains the ideal limit isolating the two classes so that their consolidated spread is insignificant or maximal.At last the yield is initialised contrasting with its mean quality.

III. FLOWCHART

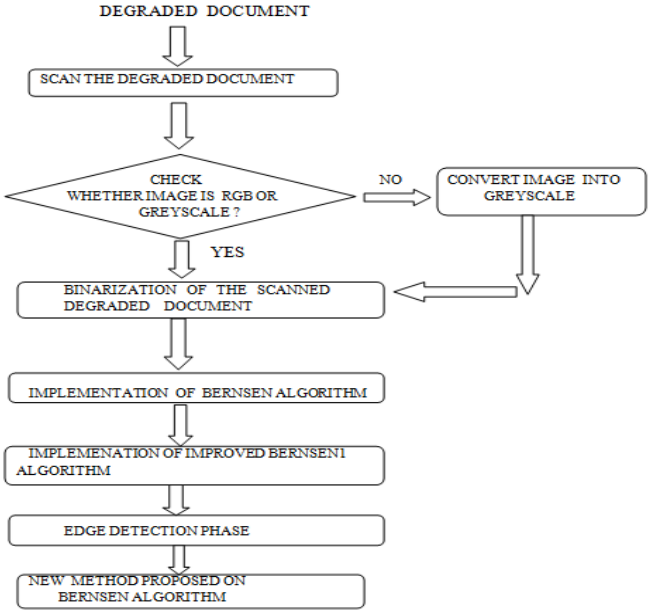


Fig 1: Flowchart For the Methodology

IV. EVALUATION MEASURES

1. PEAK SIGNAL TO NOISE RATIO: PSNR is a building term for the proportion between the greatest conceivable force of a sign and the force of undermining commotion that influences the loyalty of its representation. Since numerous signs have a wide element range, PSNR is generally communicated regarding the logarithmic decibel scale.
2. PRECISION: In example acknowledgment and data recovery with paired grouping, exactness (likewise called positive prescient quality) is the division of recovered cases that are relevant.high accuracy implies that a calculation returned significantly more pertinent results than irrelevant,Precision is the likelihood that an (arbitrarily chose) recovered report is applicable.
3. RECALL: while review (otherwise called affectability) is the division of pertinent occasions that are retrieved.high review implies that a calculation returned the vast majority of the important results.Recall is the likelihood that a (haphazardly chose) significant archive is recovered in a hunt.
4. F-MEASURE: A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

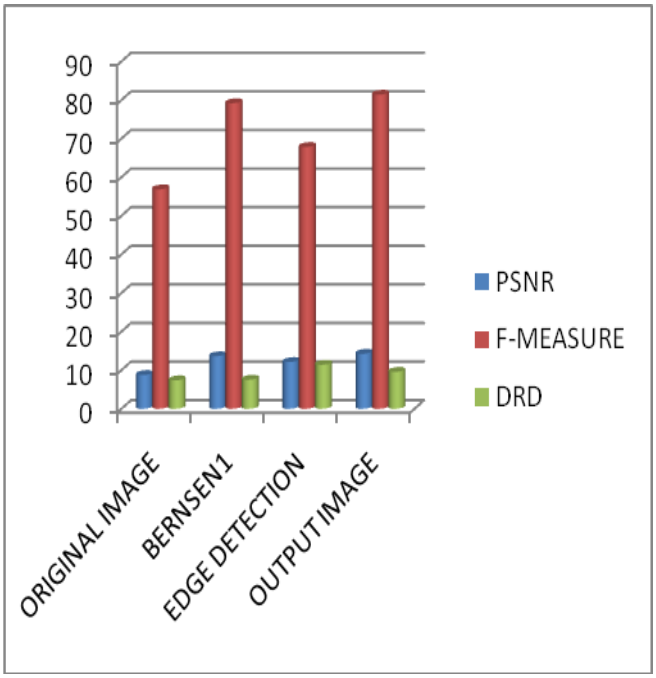
$$F = 2 * \frac{Precision * Recall}{Precision + Recall}$$

5. DISTANCE RECIPROCAL DISTORTION(DRD): DRD is utilized to quantify the mutilation in double record images.This measure correaltes with human visual preception appropriately and measures the bending for all the S flipped pixels as per the accompanying recipe

V. RESULTS & GRAPHS

EVALUATION MEASURES / METHOD	PSNR	F-MEASURE	DRD
ORIGNAL IMAGE	8.7712	56.8782	7.384
BERNSEN1	13.623	79.1854	7.5362
EDGE DETECTION	12.1198	67.8356	11.4074
OUTPUT IMAGE	14.246	81.388	9.571

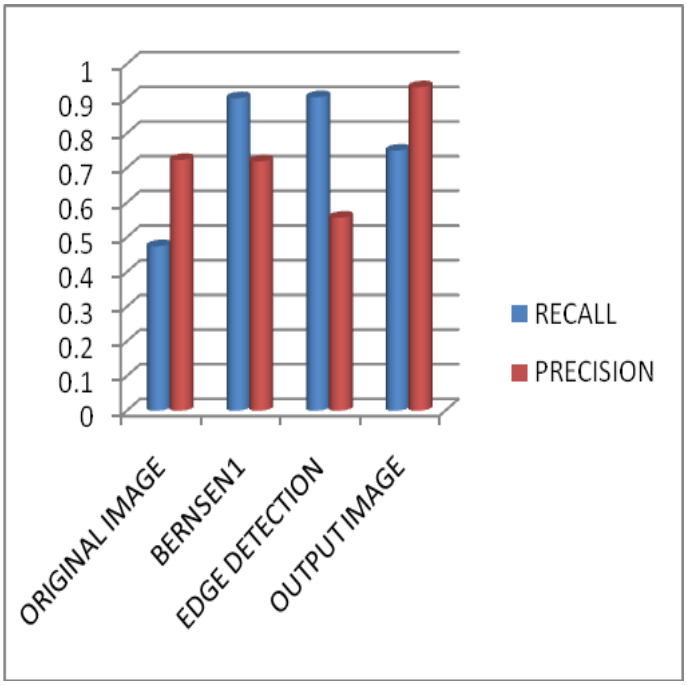
Table 1: Comparison among the Evaluation Metrics and Various Methods like PSNR, F-Measure, DRD



Graph 1: Graph Showing The Pictorial Representation Of The Evaluation Measures Like PSNR, F-MEASURE, DRD.

EVALUATION MEASURES ⇨ / METHOD ⇩	PRECISION	RECALL
ORIGINAL IMAGE	0.4744	0.7232
BERNSEN 1	0.9014	0.7194
EDGE DETECTION	0.9044	0.5568
OUTPUT IMAGE	0.7502	0.933

Table 2: Comparison among the Evaluation Metrics and Various Methods like PRECSION & RECALL



Graph 2: Graph Showing The Pictorial Representation Of The Evaluation Measures Like PRECISION & RECALL.

VI. CONCLUSION

The proposed calculation went over test image databases having debased record images. The calculation displays a versatile image difference based report image binarization procedure that is tolerant to assortment of record debasement, for example, uneven enlightenment and archive smear. The proposed strategy is simple and hearty, just couple of parameters are included. It works for various types of debased archive images. It makes utilization of the nearby image differentiate that is assessed in view of the neighborhood most extreme and minimum.The Canny's edge recognition calculation performs superior to every one of these administrators under all scenarios.The proposed technique has been tried on the different datasets. Trials demonstrate that the proposed technique outflanks most reported archive binarization routines in term of the F-measure,Precision, Recall,PSNR and DRD. The outcomes indicate particularly great adjustment into distinctive imperfection sorts, for example, brightening, commotion and determination changes. The calculation indicated strong conduct in most, even serious, circumstances in corruption and performed well against the correlation methods.

VII. FUTURE WORK

The proposed system could be enhanced to dispose of expansive clamor from foldings or page parts without tossing significant content data. Keeping in mind the end goal to make the proposed binarization calculation more successful and build the clarity of content we attempt to identify the hazardous zone in the report image in light of certain identification condition as future work. In other hand we can apply post preparing venture to calculation which will expands the execution and nature of general record image.

REFERENCES

[1] Joseph Shu, Jamie Li, and Andrei Pascovici," Multiple Layer Screening for Reducing moire Patterning and Ink Bleeding",IEEE, 0-8186-8183-7/97, 1997
[2] Neil F. Johnson and Sushil Jajodia," Exploring Steganography:Seeing the Unseen",IEEE,0018-9162/98, 1998
[3] J.Sauvola and M. PietikaKinen," Adaptive document image binarization", Pattern Recognition,Elsevier,0031-3203/99, 1999.
[4] R. D. Lins and J. M. M. da Silva," Assessing Algorithms to Remove Back-to-Front Interference in Documents",IEEE, 85-89748-04-9/06, 2006.
[5] Fadoua Drira," Towards resctoring Historic Documents Degraded Over Time", IEEE,0-7695-2531-8/06, 2006.
[6] Drira Fadoua, Frank Le Bourgcois and Hubert Emptoz," Restoring Ink-Bleed Through Degraded Document Images Using a Recursive Unsupervised Classification Technique",Springer, 2006.
[7] Laurence Likformen-Sulem, Abderrazak Zahour and Bruno Teconet," Text line segmentation of Historical documents: a survey", Springer,DOI 10.1007/s 10032-006-0023-z, 2006.
[8] Jing-Ming Guo, Member, IEEE, and Yun-Fu Liu," License Plate Localization and Character Segmentation With Feedback Self-Learning and Hybrid Binarization Techniques,IEEE transactions on vehicular technology, vol. 57, no. 3, 0018-9545, may 2008.
[9] John Ashley Burgoyne, Johanna Devaney, Laurent Pugin and Ichiro Fujinaga," Enhanced Bleed through Correction For Early Music Documents With Recto-Verso Registration",ISMIR 2008.
[10] Oliver Nina, Bryan Morse and William Barrett," A Recursive Otsu Thresholding Method for Scanned Document Binarization",IEEE,978-1-4244-9497-2/10, 2010.
[11] Tarnjot Kaur Gill," Document Image Binarization Techniques- A Review", International Journal of Computer Applications (0975 – 8887) Volume 98– No.12, July 2014.
[12] Xiaojun Wang," Recovery of Blurring Scanned Manuscript Image Based on Wavelets Transform Algorithm",IEEE, 2010 3rd International Congress on Image and Signal Processing (CISP2010),978-1-4244-6516-3/10, 2010.
[13] G.F. P e Silva, R. D. Lins, J.M. Silva, S. Banergee, A. Kuchibhotla and M. Thielo," Enhancing the Filtering-out of the Back-to-Front Interference in Color Documents with a Neural Classifier",IEEE,2010 International Conference on Pattern Recognition,1051-4651/10, 2010.

[14] Oliver Nina, Bryan Morse, and William Barrett," A Recursive Otsu Thresholding Method for Scanned Document Binarization",IEEE,978-1-4244-9497-2/10, 2010.
[15] Jian-Xia Wang, Wan-Zhen Zhou, Jing-Fu Xue And Xin-Xin Liu," The Research And Realization Of Vehicle License Plate Character Segmentation And Recognition Technology",IEEE,Proceedings of the 2010 International Conference on Wavelet Analysis and Pattern Recognition, 978-1-4244-6531-6/10, 2010.
[16] Raman Jain , Volkmar Frinken , C.V. Jawahar and R. Manmatha," BLSTM Neural Network basedWord Retrieval for Hindi Documents", International Conference on Document Analysis and Recognition,2011.
[17] Roisin Rowley-Brooke and Anil Kokaram," Bleed-through removal in degraded documents",Irish Research Council for Science, Engineering, and Technology (IRCSET),2011.
[18] Rajbir Kaur and Amandeep Kaur," An Effective Algorithm for Ink-Bleed through Removal in Document Images",IJCST Vol. 2, Issue 4,ISSN : 0976-8491,Oct.-Nov. 2011.
[19] Hafizan Mat Som, Jasni Mohamad Zain and Amzari Jihadi Ghazali," Application Of Threshold Techniques For Readability Improvement Of Jawi Historical Manuscript Images",Advanced Computing: An International Journal (ACIJ), Vol.2, No.2, March 2011.
[20] Rabeux Vincent, Journet Nicholas and Domenger Jean Philippe,"Document recto-verso registration using a dynamic time warping algorithm", International Conference on Document Analysis and Recognition,2011.
[21] Noureddine El Makhfi, Omar El Bannay, Rachid Benslimane and Noureddine Rais," System of Indexing, Annotation and Search in the old Arabic Manuscripts",IEEE,978-1-4673-0115-2/11, 2011.
[22] Bolan Su, Shijian Lu and Chew Lim Tan," A Robust Document Image Binarization Technique for Degraded Document Images",IEEE Transaction On Image Processing, 2012.
[23] Angelika Garz, Andreas Fischer, Robert Sablatnig and Horst Bunke," Binarization-Free Text Line Segmentation for Historical Documents Based on Interest Point Clustering", 10th IAPR International Workshop on Document Analysis Systems, 2012.
[24] Keiichiro Shirai, Masashi Akita, Masayuki Okamoto , Kazuya Tanikawa, Takaaki Akiyama and Tetsuji Sakaguchi,"Removal of Background Patterns and Signatures for Magnetic Ink Character Recognition of Checks",IEEE,10th IAPR International Workshop on Document Analysis Systems, 978-0-7695-4661-2/12, 2012.
[25] Naveen Sankaran T and C V Jawahar," Recognition of Printed Devanagari Text Using BLSTM Neural Network",ICPR(International Conference on Pattern Recognition),Report No: IIIT/TR/2012/-1, 2012.
[26] C.A.B.Mello, D.C.Costa and T.J.dos Santos," Automatic Image Segmentation of Old Topographic Maps and Floor Plans",IEEE International Conference on Systems, Man, and Cybernetics, 978-1-4673-1714-6/12, October 14-17, 2012.
[27] D. Fernandez, J. Llados & A. Fornes and R. Manmatha," On Influence of Line Segmentation in EfficientWord Segmentation in Old Manuscripts",IEEE,International Conference on Frontiers in Handwriting Recognition,978-0-7695-4774-9/12, 2012.
[28] Sukhjinder Singh, R. K. Bansal and Savina Bansal," Comparative Study and Implementation of Image Processing Techniques Using MATLAB", International Journal of Advanced Research in Computer Science and Software Engineering,Volume 2, Issue 3, ISSN: 2277 128X, March 2012.
[29] K. Ntirogiannis , B. Gatos and I. Pratikakis," A combined approach for the binarization of handwritten document images",Elsevier B.V,2012.
[30] Mohamed Cheriet, Reza Farrahi Moghaddam and Rachid Hedjam," A learning framework for the optimization and automation of document binarization methods",Elsevier,November 20, 2012
[31] Sukhpreet Singh," Optical Character Recognition Techniques: A Survey", Journal of Emerging Trends in Computing and Information Sciences, Vol. 4, ISSN 2079-8407, No. 6 June 2013.
[32] Naveen Sankaran T, Aman Neelappa and C V Jawahar," Devanagari Text Recognition: A Transcription Based Formulation",12th International Conference on Document Analysis and Recognition, 25-28 Aug. 2013.
[33] Sergey Milyaev, Olga Barinova, Tatiana Novikova, Pushmeet Kohli and Victor Lempitsky," Image binarization for end-to-end text understanding in natural images",IEEE,12th International Conference on Document Analysis and Recognition1520-5363/13, 2013
[34] Shrikant Baronia and Anoop Namboodiri," Ink-Bleed Reduction using Layer Separation",IEEE,12th International Conference on Document Analysis and Recognition,1520-5363/13, 2013.
[35] Mr. Sadanand S. Muktawar And Miss.Juhi R. Bobade," Advancement in Image processing using Android", International Journal of Advanced Research in Computer Science,Volume 4, No. 6, ISSN NO- 0976-5697, May 2013.
[36] Konstantinos Ntirogiannis, Basilis Gatos, and Ioannis Pratikakis," Performance Evaluation Methodology for Historical Document Image Binarization",IEEE Transactions On Image Processing, Vol. 22, No. 2, February 2013.
[37] Mohamed Aymen, Charrada Najoua and Essoukri Ben Amara," Development of a database with ground truth for old documents analysis", IEEE, 10th International Multi-Conference on Systems, Signals & Devices (SSD),978-1-4673-6457-7/13, 2013.
[38] Bolan Su, Shijian Lu, and Chew Lim Tan," Robust Document Image Binarization Technique for Degraded Document Images",IEEE Transactions On Image Processing, Vol. 22, No. 4, April 2013
[39] Sandeep Tiwari, Shivangi Mishra, Priyank Bhatia and Praveen Km. Yadav," Optical Character Recognition using MATLAB",International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), ISSN: 2278 – 909X,Volume 2, Issue 5, May 2013.
[40] Mauricio Villegas and Alejandro H. Toselli," Bleed-through Removal by Learning a Discriminative Color Channel", International conference on frontiers in handwriting recognition,ICFHR 2014, September 2, 2014.
[41] Jensiya Asirvatham, Akila Agnes," Survey on Restoration of Degraded Document using Different Thresholding Techniques",International Journal Of Engineering Sciences & Research Technology,ISSN: 2277-9655,April, 2014.
[42] Tarnjot Kaur Gill," Document Image Binarization Techniques- A Review", International Journal of Computer Applications (0975 – 8887) Volume 98– No.12, July 2014.
[43] Geetanjali Thakur," A Comprehensive Review On Analysis Of Image Binarization For Degraded Documents",International Journal of Advance Research In Science And Engineering IJARSE, Vol. No.3, ISSN-2319-8354, Issue No.7, July 2014.
[44] Abdenour Sehad, Youcef Chibani and Mohamed Cheriet," Gabor Filters for Degraded Document Image

Binarization",IEEE,14th International Conference on Frontiers in Handwriting Recognition,2167-6445/14, 2014.

[45] Haneen Khader, Abeer AI-Marridi, Hena Alpona, Suchithra Kunhoth, Abdulaali Hassaine, and Somaya Al-maadeed," An Interactive Annotation Tool for Indexing Historical Manuscripts",IEEE,978-1-4799-2806-4/14, 2014

[46] Jingning Ji, Liangrui Peng and Bohan Li," Graph Model Optimization based Historical Chinese Character Segmentation Method",IEEE,11th IAPR International Workshop on Document Analysis Systems,978-1-4799-3243-6/14, 2014.

[47] N. Chaki," A Comprehensive Survey on Image Binarization Techniques", Exploring Image Binarization Techniques, Studies in Computational Intelligence 560, Springer, DOI: 10.1007/978-81-322-1907-1_2, 2014.

[48] Indrajeet Kumar, Jyoti Rawat and Dr. H.S. Bhadauria," A Conventional Study Of Edge Detection Technique In Digital Image Processing", International Journal of Computer Science and Mobile Computing IJCSMC, Vol. 3, Issue. 4, ISSN 2320–088X, pg.328 – 334, April 2014.

[49] Rinku Kalotra , Sh. Anil Sagar," A Novel Algorithm for Blurred Image Restoration in the field of Medical Imaging",International Journal for Science and Emerging Technologies with Latest Trends, 17(1): 21- 26(2014),ISSN No. (Online):2250-3641, 2014

[50] Biswa Ranjan Mohapatra, Ansuman Mishra and Sarat Kumar Rout," A Comprehensive Review on Image Restoration Techniques",International Journal of Research in Advent Technology, Vol.2, No.3, E-ISSN: 2321-9637, March 2014

[51] Ravneet Kaur, Er. Navdeep Singh," Image Restoration - A Survey",IOSR Journal of Computer Engineering (IOSR-JCE),e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 4, Ver. III, PP 107-111, Jul – Aug. 2014.

[52] Sayali Shukla, Ashwini Sonawane, Vrushali Topale and Pooja Tiwari," Improving Degraded Document Images Using Binarization Technique",International Journal Of Scientific & Technology Research Volume 3, Issue 5, ISSN 2277-8616, May 2014.

[53] Anamika Maurya, Rajinder Tiwari," A Novel Method of Image Restoration by using Different Types of Filtering Techniques", International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 3, Issue 4, ISSN: 2319-5967 ISO 9001:2008 Certified. July 2014.