ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Distorted Heavily Printed Document Recognition Using Simulink - Edge Thinning

Dr.K.P.Kaliyamurthie[1] R. Indra Gandhi[2] , Sundararajan.M, Arulselvi S

2 PG Student, Department of Computer Science and Engineering, Bharath University, Chennai, India. Email:

1 Professor and Head, Department of Computer Science and Engineering, Bharath University,

Chennai, India.

Director, Research Center for Computing and Communication, Bharath University, Chennai, Tamil Nadu, India

Co-Director, Research Center for Computing and Communication, Bharath University, Tamil Nadu, India

ABSTRACT: Many techniques out there for character recognition and lots of authors have contributed their work on character recognition permanently quality document pictures. As way as distorted documents ar involved, solely a fewer tries were created that too for Roman scripts. Most of our valuable aged records have distortion. it would be as a result of poor maintenance, poor paper quality, blurring of ink and damages done by bookworms. there's a requirement to own several of those documents hold on in electronic format to reinforce their life and readability. generally the conversion conjointly introduces distortions to the pictures. This paper is a shot to spot major drawback in distorted text pictures and to implement a similar exploitation Simulink – Edge dilution technique to observe edges. the most goal of edge detection is to find Associate in Nursingd establish sharp discontinuities from a picture that helps to search out the perimeters in an input image whereas eliminating the low distinction boundaries.

KEYWORDS: heavily printed; distored, edge detection, thinning

I. INTRODUCTION

Edge detection is one of the fundamental steps in image processing, image analysis, image pattern recognition, and computer vision techniques. In recent past pattern recognition directly deals computer vision systems, orientation and intensity information about edges as primary input for further processing to document identification. The edge can have different meaning in various contexts. Accordingly, different edge detection algorithms can recognize edges in different forms of representation and each of them can be considered as a genuine edge detection algorithm.

Edge thinning is a technique used to remove the unwanted spurious points and filters out unwanted or insignificant information in an image. This technique is employed after the image has been filtered for noise [1] to detect the edges and after the edges have been smoothed using an appropriate threshold value. This removes all the unwanted points and if applied carefully, results in one pixel thick edge elements. If the edge happens to be the boundary of a region, then thinning could easily give the image parameters like perimeter without much algebra.

II. REVIEW OF LITERATURE

The edge detection process typically results in an edge map which is usually a binary image. All images describe the major classification of each pixel of the images, as well as some other edge attributes such as magnitude and orientation. Edge detection is a low level operation used in image processing and computer vision applications. The main goal of edge detection is to locate and identify sharpdiscontinuities from an image. These discontinuities are due to abrupt changes in pixel intensity which characterizes boundaries of objects in a scene [2].

The main goal of edge detection is to locate and identify sharp discontinuities from an image. These discontinuities are due to abrupt changes in pixel intensity which characterizes boundaries of objects in a scene. Edges give boundaries between different regions in the image. These boundaries are used to identify objects for segmentation and matching purpose [3]. The main of image analysis is to extract meaningful features from image data in order to reduce

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

computational processing cost in higher level processes [4]. Image analysis can be considered as a data reduction process and its operations usually focus on reducing image data.
In First order derivative the input image is convolved by an adapted mask to generate a gradient image in which edges are detected by threshold. In second order derivative, these are based on the extraction of zero crossing points which indicates the presence of maxima in the image [5]. Since the second order derivative is very sensible to noise, and the filtering function is very important. These operators are derived from the Laplacian of a Gaussian (LOG), and proposed by Marr and Hildreth, in this, the image is smoothed by a Gaussian filter. For this operator we have to fix some parameters such as the variance of the Gaussian filter and thresholds.
Some methods are available for their automatic computation [7], but in most cases their values have to be fixed by the user. A significant problem of LoG is that the localization of edges with an asymmetric profile by zero-crossing points introduces a bias which increases with the smoothing effect of filtering [8].

An interesting solution to this problem was proposed by Canny [9] and Shen [10], which says in an optimal operator for step edge detection, includes three criteria: good detection, good localization, and only one response to a single edge.

III. MAJOR PROBLEM IN DISTORTED TEXT IMAGE

Following are some well-known reasons in distortions of document images that play a fair role in increasing the difficulty in recognition [11].

A. Distortion by Natural calamities

Due to poor maintenance, ageing of the document (old), damages done by bookworms like silverfish and booklouse, disconnection of arbitrary direction due to the presence of foreign material etc., are listed out in this category

B. Distortion in the Paper

The distortions that come in this category include, Physical distortions like Vertical cuts, hair line holes, coffee stains, discolor, yellowing, wrinkles, damaged typeface, point size, spacing and typeset imperfections which are caused by slippage can show variations in pressure and position.

C. Distortion during Printing

Floating inks from facing pages, heavy printing inks spread showing large ink blobs, bleeding and scattering results in merging the disjoint characters or components. Low quality of printing ink usage, light impression on paper, unnecessary toner dropout, baseline variations, strikethrough effects and paper defects can be considered as the source in this category.

D. Distortions during Scanning

Skewness (geometric deformation), wrong threshold, resolution reduction, blur, sensor sensitivity noise, sampling, defocusing, low contrast, non-uniform illumination and non-rectilinear camera positioning are the distortions that can be positioned in this category.

E. Distortions During photocopiers and fax machines

Low quality photocopies, taking photocopy of the already digitized photocopier (not from the original one), skew, streaking, shading, noise effect from electronics components, general copier “grunge” caused by a dirty glass or background can be considered in this category.

Distorted documents do not include all the ideal properties of a document Li et al. [12] have discussed various defect models, their applications and methods for validating document defect models. Over the last few years, a lot of importance is being given to the problem of modeling document image defects so that a formal evaluation of the different OCR systems can be done.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

The methods and algorithms used in the development of these OCRs are often biased by the researcher's choice of the training and the test data sets. As a result, such systems give excellent performance for the data sets chosen by the researcher.

In many cases, however, the recognition accuracy falls sharply when even a slightly distorted image is chosen.

The fall in recognition accuracy is often high compared to the visual nature of the distortion, i.e., the distortion as perceived by the human eye. It has, therefore, been felt necessary to model the defects quantitatively and experiment with extensive simulation to determine the nature of image defects that result in higher failure rate.

There exist different steps for recognition of distorted documents. For each distortion type, we devise a unique system with its own set of steps.

IV. IMPLEMENTATION

The Step by Step implementation procedure using the edge detection - Simulink block, for given distorted document images to detect resultant edges help us to explain each design in detail.

Step 1: The Edge Detection Block enable to Simulink library Browser to process Video and Image Processing Blockset. (Fig.1)

Step 2 : Select Parameters Method as "Sobel"

Step 2.1 : Choose output mode as binary.

Step 2.1.1: Check threshold value as user define factor

Step 2.1.2: Check Edge thinning to thin the unwanted edges.

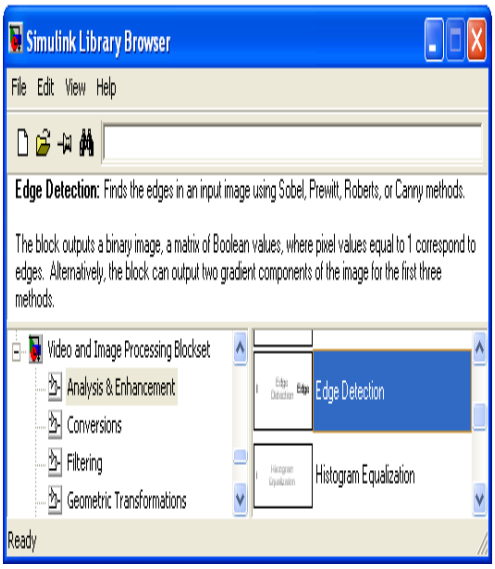

Fig.1. Edge Detection Block of Simulink Library

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

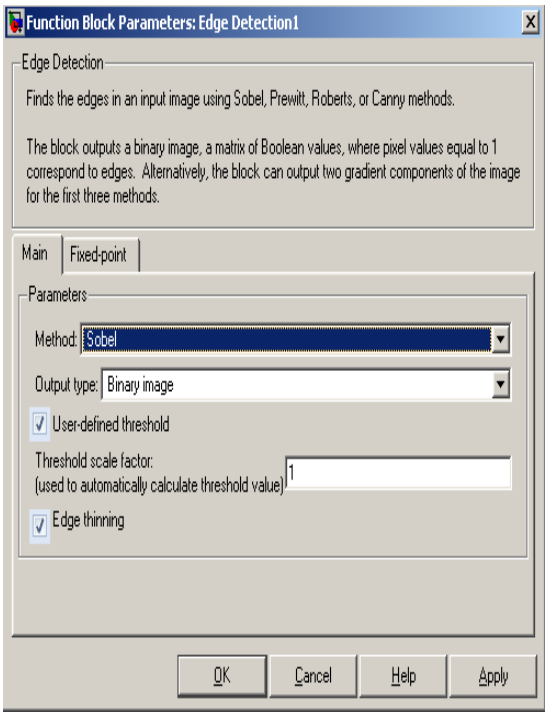

Fig.2. Threshold Scale factor Detector for Edge Thinning using Sobel

Step 3: Construct various connection action blocks using Video viewer to find edges (Fig.3)

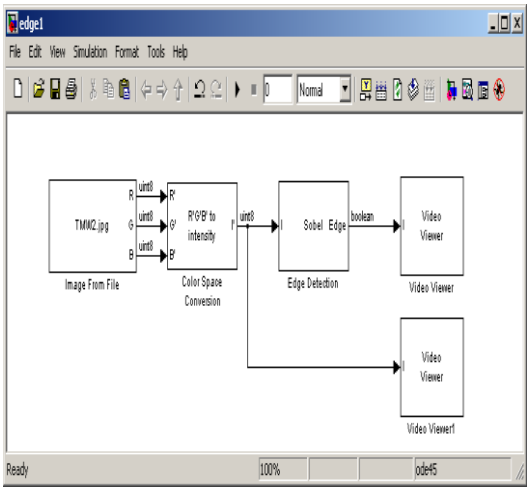

Fig. 3. Final Edge Detection Simulation Model using Sobel Method

Step 4: Construct a block to load input file with the property block name of "Image from File".

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

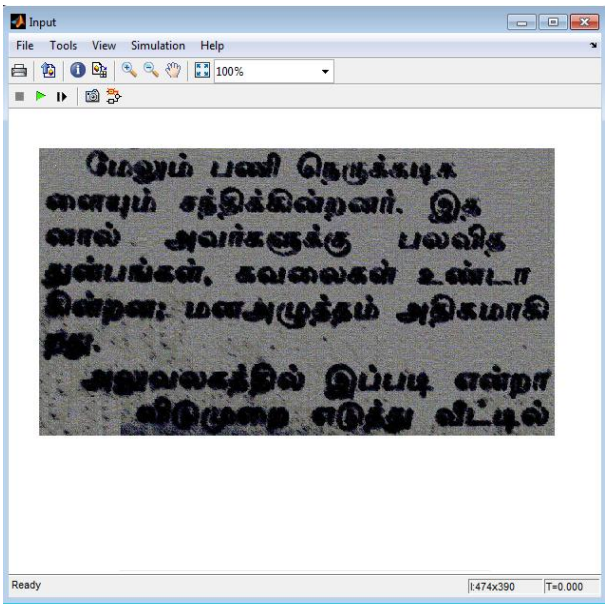

Fig. 4. Heavily Printed Distorted Image File Loaded in Simulation

Step 5: Linking input image happens in this block which executes Color space conversion (Fig 5)
Step 5.1: Image data type conversion happens at this stage.
Step 6: Selection of Sobel Edge detection takes place
Step 7: Boolean operation result to output file.
Step 8: Double click generate code & create project operation to create respective coding for the blocks.

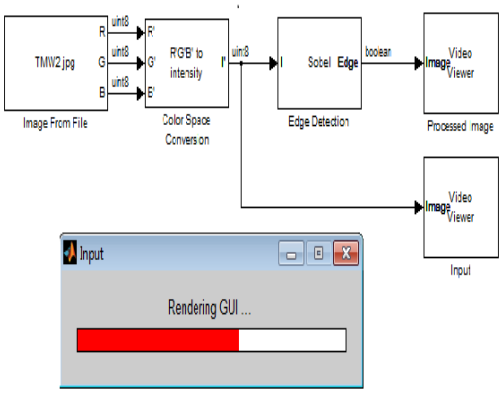

Fig. 5. Rendering GUI Model for Distorted Input Image

Step 9: Build the project.
Step 10: By Double-click loading and running option results will be displayed in the screen

Execution of Simulink-Sobel model results three different operational dialogue windows. Sample heavily printed input image is loaded in "Image from file" as mentioned in Fig 3 with the name "TMW2" as shown in Fig 4. Processed image dialogue windows evident for edge detection as noted in Fig 6. Expected reconstructed heavily printed distorted document after gray scale processing is shown in Fig 7. A gray scale image conversion adds value to resultant document.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

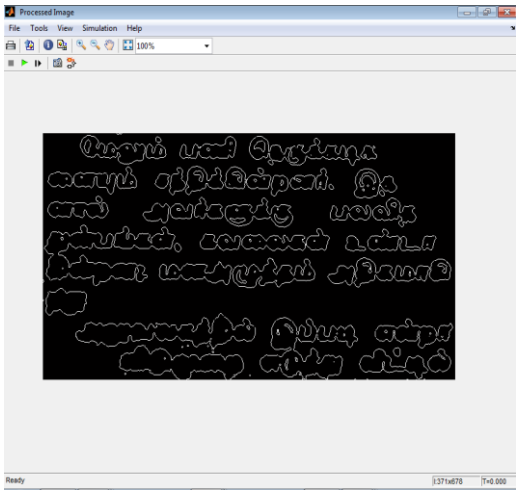

Fig.6. Edge Thinning Document Image after Edge Detection processing

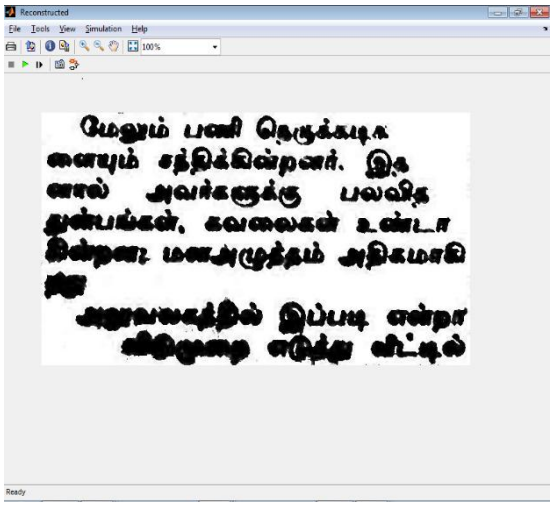

Fig.7. Reconstructed Heavily Printed Distorted Image after Gray Scale Processing

V. CONCLUSION

The proposed method proves to perform well regardless of the intensity differences between foreground and background. Especially we have used edge thinning over heavily printed characters to satisfy the fundamental requirement of thinning know as quality of thinned result. This process can remove irregularities in letters and in turn, make the recognition simpler because they only have to operate on one pixel wide. It also reduces the memory space required for storing the information about the input characters and no doubt, this process reduces the processing time too. This can be used in almost all the types of distortions to identify the original character structure. The resultant Reconstructed Heavily Printed Distorted Image after Gray Scale Processing evident the same.

REFERENCES

1. R.C.Gonzalez and R.E.Woods, "Digital Image Proccessing", Prentice Hall, 2007.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

2. Vijayaprakash S., Langeswaran K., Gowtham Kumar S., Revathy R., Balasubramanian M.P., "Nephro-protective significance of kaempferol on mercuric chloride induced toxicity in Wistar albino rats", Biomedicine and Aging Pathology, ISSN : 2210-5220, 3(3) (2013) pp.119-124.
3. Saket Bhardwaj, Ajay Mittal, "A Survey on Various Edge Detector Techniques", Elsevier LtdProcedia Technology 4, Pgno: 220 – 226, 2012.
4. Sundar Raj M., Arkin V.H., Adalarasu, Jagannath M., "Nanocomposites based on polymer and hydroxyapatite for drug delivery application", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S5) (2013) pp.4653-4658.
5. H. Voorhees and T. Poggio, " Detecting textons and texture boundries in natural images" ICCV 87:250-25,198.
6. Kumar S., Das M.P., Jeyanthi Rebecca L., Sharmila S., "Isolation and identification of LDPE degrading fungi from municipal solid waste", Journal of Chemical and Pharmaceutical Research, ISSN : 0975 – 7384 5(3) (2013) pp.78-81.
7. S.E.Umbaugh, "Computer Imaging : Digital Image Analysis and Processing", CRC Press, 2005
8. Laljee R.P., Muddaiah S., Salagundi B., Cariappa P.M., Indra A.S., Sanjay V., Ramanathan A., "Interferon stimulated gene - ISG15 is a potential diagnostic biomarker in oral squamous cell carcinomas", Asian Pacific Journal of Cancer Prevention, ISSN : 1513-7368, 14(2) (2013) pp.1147-1150.
9. M. Juneja,P sandhu, "Performance evaluation of edge detection Techniques for images in spatial domain" International Journal of Computer theory and Engineering, Vol. 1, No. 5, Dec, 2009 1793- 8201.
10. Subha Palaneeswari M., Ganesh M., Karthikeyan T., Manjula Devi A.J., Mythili S.V., "Hepcidin-minireview", Journal of Clinical and Diagnostic Research, ISSN : 0973 - 709X, 7(8) (2013) pp.1767-1771.
11. Marr,hildreth., "Theory of edgedetection.", Proc.R. Soc. Lond.(1980), B 207, pp. 187-217.
12. Dellepiane, Giusto, Serpico, Vernazza, "Automatic parameter computation for edge detection by the zero crossing method. 12 e Colloque GRETSI sur le traitement du signal et des images",Juan-les-Pins (12-16 juin 1989), pp.617-612.
13. J F Canny, "A computational approach to edge detection." 1EEE Trans. PAMI (1986), 8, n ~ 6, pp. 679-698.
14. Deriche . R, " Fast algorithms for low-level vision", IEEE Trans. PAMI (1990), 12, n ~ 1, pp. 78-87.
15. Shen J, Castan S, "An optimal linear operator for step edge detection", CVGIP: Graphical Models and Image Processing (1992), 54, n ~ 2, pp. 112-133.
16. R. Indra Gandhi, K. Iyakutti and C. Jothi Venkateswaran, "An analysis of various types of Distortions of Tamil Scripts", World Tamil Internet Conference- Coimbatore 2010, June 23-27,pp362-371, 2010.
17. Y. Li, D. Lopresti, G. Nagy and A. Tomkins, "Validation of image defect models for optical character recognition", *IEEE Transactions on PAMI*, Vol. 18(2), pp. 99-108, 1996.
18. Sangeetha Rajagurusamy, Analysis of Work study in An Automobile Company ,International Journal of Innovative Research in Science, Engineering and Technology, ISSN: 2319-8753 , pp 5622-5631,Vol. 2, Issue 10, October 2013.
19. V.G.Vijaya,Analysis of Rigid Flange Couplings ,International Journal of Innovative Research in Science, Engineering and Technology ,ISSN: 2319-8753 , pp 7118-7126,Vol. 2, Issue 12, December 2013.
20. V.G.Vijaya ,DESIGN OF HUMAN ASSIST SYSTEM FOR COMMUNICATION ,International Journal of P2P Network Trends and Technology(IJPTT),ISSN: 2319-8753 ,pp 3687-3693,Vol. 2, Issue 8, August 2013.
21. V.G.Vijaya, V.Prabhakaran ,Design of Human Assist System for Communication ,International Journal of Innovative Research in Science, Engineering and Technology ,ISSN: 2249-2651,pp 30-35, Volume1 Issue3 Number1–Nov2011.
22. V.Krishnasamy, R.Kalpana devi ,Isomorphous Salts with Abnormal Water Of Hydration ,International Journal of Innovative Research in Science, Engineering and Technology,ISSN: 2319-8753 , pp 3500-3509,Vol. 2, Issue 8, August 2013.
23. Veera Amudhan R ,Tracking People in Indoor Environments ,International Journal of Innovative Research in Science, Engineering and Technology,ISSN: 2319-8753 , pp 15996-16003,Vol. 3, Issue 9, September 2014.