# A New and Efficient Technique to remove Back to Front Interference in Historical Document Images

Bhuvnesh Malik<sup>1</sup>, Rajiv Kumar<sup>2</sup>

<sup>1</sup>M.Tech., CSA, Computer Science and Engineering Department, Thapar University <sup>2</sup>Assistant Professor, Computer Science and Engineering Department, Thapar University



## Introduction

- Historical documents are of great importance to us due to their cultural and scientific value.
- Digital preservation is the foremost requirement of today to preserve the historical documents.
- Specialized processing techniques are required to improve reading ability and remove noise from these digitized historical document
- Once these documents are digitized, they can be made available in digital libraries or on the internet for wider dispersion.
- We proposed a new technique, which removes the noise and improves the reading ability of stained historical document images.



# **Proposed Algorithm**

#### Pseudo-code:

Input: A grayscale image

- Background removal is performed by using median filter.
- Adjust image intensity values (increasing contrast) of image resulted from step 1.

$$J = image_{adjust}(I).$$

3. A global threshold (T) applied to this image (J).

Where 
$$T = average(J)$$
,  
 $If J(m, n) > T : Out(m, n) = 255$   
 $ElseOut(m, n) = 0$   
 $End$ 

4. Apply morphological erosion operation by using a disk of radius 1 on the binary image (Out), we get after global threshold.

$$Im_{morph} = image_{erosion}(Out, disk(1)).$$

5. Add images J and  $Im_{morph}$ .  $Im_{add} = J + Im_{morph}$ .

6. Apply adaptive threshold method on the intermediate image  $Im_{add}$ .

# Step by step changes in historical document

THE THE PROPERTY OF THE PARTY O STORES OF STREET AND A THORE AND STREET AND CAPTE PRESENT CALANTE PARTY ONL The state of state and the species and part with THE MENT OF THE PARTY OF HE WAN UT a contact pear there dept these into the present HAPTETO MEAN AND HOSE OF THE OTH

ATTHEM MOMBANTALINOI & SELECT

THE PARK OF MANAGES AS IN WASHING WITH

Trub & Abstratel a style A Tris one

A C OF MA GOAL MEN THE A STREET A STREET AND A WALL

THE OWNER OF STREET STREET, ST

OVER WEST STORY OF STREET STREET ST.

CONTROL OF A PROPERTY OF A PICKET OF A PARTY OF A PARTY

TYPES OF MY MOST NA MONE AFTER KEY BHITTH A JOH I STEMA TO TO JE AND PEND TO LA LA LONG ALAND AND AND AND AND SHAPE

E-T IT FOR MAP TEPIANADT OF DIMATE

TENHA DONA MATERIA MATERIA

Pere Kaleria Kal somete Lon

HAPPETE HEAR ALTHOUGH COTTON OF

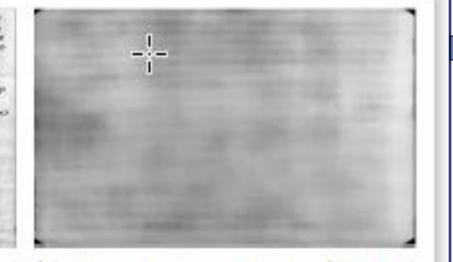
THE OF PERSONS AND THE PROPERTY.

CAPPARTIMETAL FURNISHED MAN OF

a diplan post time Joy's necessary to a

HAPPETE HEXINAL HONCETTIN ON

SHOULD BE STANKED TO SELECT STANKED ST



#### (a) Grayscale Image

(b) Image after Applying Median Filter (background extraction)

AND THE PROPERTY OF THE PROPERTY OF A PARTY TWHEN TO WELL BY TO THE PARTY OF THE PARTY O OI CYTTE-PICEARYOVALAPT VITE MALAP OI GALLET CLUL TOATEN LALL MAND TYPIAC OF OVER COTTON & CHOME AND BY COST OF A SECTION AS A PRINCIPLE OF WALLES SEPT ONTO MAJOR COVERNY LOS M. SCALLPRING ALAP TEPENTERPIENDE CARETTE TALL ONE e-c-14 pith erich in Address a. a.t. oversian SEN HA GON KATHOT Y TAPE YALLE THE OWNERS WAS TRUBERS OF WE AS ME SI THE ST TO THE PART OF MARKET SA PROPERTY CATHARTING THE STORY PART MEN OF mercan pear to the period of the pear HAPITO HEAD AS HONCESTIN ON the spine or police | And to rest had a

#### (c) Image after Removal of Background (d) Image after Adjusting Intensity of Pixels

THE TYPINGS SAMETEY AND SHOW SELL SAN WAS A ASSESSED SAN ON WHEEL SAN ++MMADONKAMETYHAPWYMIT Treatment at tiopered Heaved 

(e) Image after Applying Global Binariza-

OF PLANTEL CARA HEEMET SANTAGE STORE CHARY YES SITE OF STHEM TWHEN HONOTHAT TWO | PAFIER! OF THE PICEATTOTALAP TYPIC HALAP TETTAL OF OF MESTINA SHOULD AGE PLI OHICK ATH HENRY TOTA MANTEL ALAP TYPEN HOTE CALKYTE TAXIS OHE ET IT FOR MANY TYPIANARY BY BLEATTE TO MA BOM KAINING THAT WY MIE Tre-ma bent was the beat at me we get H-TTOY STRATES Y ALCIE LA PROPERTY CATEANTINE TATE THE TATE OF Selve Karent Kit tomber e Len HAPTETT HEADIAN HONCOTTING OT LE ON TE OY MILLIANA PIEN NALIO

(f) Image after Erosion Operation

moreorpeas ware jupe amount to pres

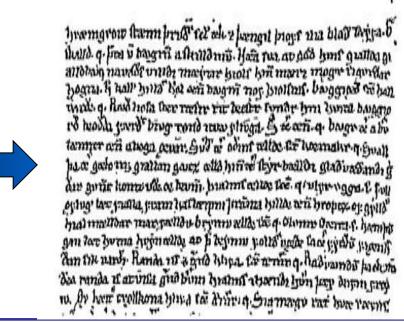
HAPTER HEAM ALMONE COTTON OF

orthographony period of the en the election of the enter ANHOR MONOTHATTHOI PAPICAL THE SHIEN VIGITAMENTAL & SCHILEM THE POST OF THE PERSON OF THE WELL HAN WAS LASTON ON CITYLE TON HA BON ICATTO TOTAL W PARTE Treatentor 147 Hopers of sie or of States & Same Lond west a by Lond His CAFECAR FINOTOJE FOR THE PAPE MEN OF the distance for an about some some such that the second HAPTELY HEALH AS HOHOSETIM THE

Image after the of (h) Binarized Image using Adaptive Thresh-IntensityAdjustedImage(d)and ErodedImage(f)

### Results

hand q fra v bagin afkindmv. Han rua av des him giantagi allobar navdet unite mæjrar hou hin marr mogre rigurdar hogia. h hall hille hot en bagne nog hielfint bargered rehan unite g had hote ter ræfir rit beste tynde hin hunte bargere re hoode gard bing rond mas plinge. E se en g barger e abs tæmper æit aloga gener. Sud af odme ælter be kormene a Gwall ha æ gæle ing grattan gauez ælte hiñ æ flyr bællde gladvædande g dar gutir kome vile se kuni huatunt setter toe q vilye wagant full dar gotte kome vie de kenn hallet ette de dige viget gotte opper og gold ogbeg lee gralle gram hallet min brille ett broper og gold hall mællder mar pællde brynne ællde lee goldene været hæmpe gan lor home home homelde av filopine polle vede lav golde stæmf lamin home hande itt e gud hiper færende hadvænde hadvænde hadvænde bedene des rande it atville gud bom hralmi itsænde hom hæy digen græ w. De han crallhona him a ta drive a Sna margo rat how raying





# Results (Comparision with existing methods)

Thresholding Technique	Evaluation N	Evaluation Metrics	
	RMS	SSIM	
Otsu	2.8348	0.5025	
Iterative method	4.3045	0.8376	
Niblack	4.1614	0.1225	
Sauvola	6.0655	0.6700	
Kittler	5.8194	0.6865	
NICK	5.4186	0.7738	
Proposed method	2.8272	0.9296	



## Conclusion

The proposed method gives highest SSIM value which being near to 1 proves that it works well. On the other hand, RMS value is the lowest which being near to 0 proved that the proposed method outperforms these existing methods both visually as well as experimentally.

# Acknowledgments

I would like to thanks Dr. Rajiv Kumar, Assistant Professor, CSED, without the guidance of who, this work would not have been possible and Dr. Maninder Singh, Head, CSED for his constant support and encouragement.

# **Bibliography**

Bhuvnesh Malik has received his B.Tech Degree in CSE from UIT, Dehradun in 2012. He is currently pursuing his M.Tech. with specialization Computer Science and Applications from Thapar University. At present, he is working in the field of preprocessing of historical document images under the supervision of Dr. Rajiv Kumar.

## **Further Information**

For more information, visit at:

https://www.youtube.com/channel/UCcEHD4vsoPKgr-oK3Co3liQ Download poster (in pdf) at:

For any queries, mailto: bhuvneshmaalik@gmail.com