A New and Efficient Technique to remove Back to Front Interference in Historical Document Images

Bhuvnesh Malik[1], Rajiv Kumar[2]

[1]M.Tech., CSA, Computer Science and Engineering Department, Thapar University
[2]Assistant Professor, Computer Science and Engineering Department, Thapar University

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- Historical documents are of great importance to us due to their cultural and scientific value.
- Digital preservation is the foremost requirement of today to preserve the historical documents.
- Specialized processing techniques are required to improve reading ability and remove noise from these digitized historical document images.
- 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- We proposed a new technique, which removes the noise and improves the reading ability of stained historical document images.

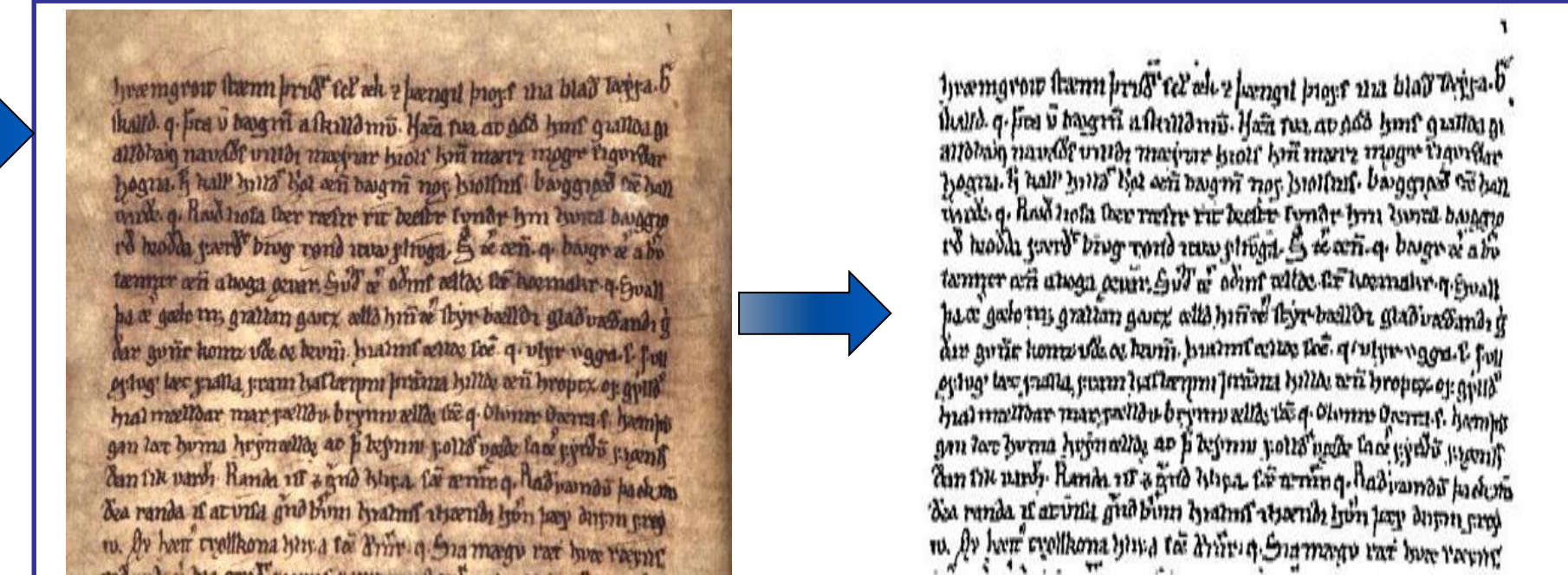## Proposed Algorithm

Pseudo-code:

Input: A grayscale image

1. Background removal is performed by using median filter.

2. 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 $T = average(J)$,
$If\ J(m,n) > T : Out(m,n) = 255$
$Else\ Out(m,n) = 0$
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Step by step changes in historical document

(a) Grayscale Image

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

(c) Image after Removal of Background

(d) Image after Adjusting Intensity of Pixels

(e) Image after Applying Global Binarization

(f) Image after Erosion Operation

(g) Image after the Addition of *Intensity Adjusted Image(d) and Eroded Image(f)*

(h) Binarized Image using Adaptive Threshold.

Results

Results (Comparision with existing methods)

Thresholding Technique	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

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*