

# Efficient Binarization for Historical Document Analysis

Florian Westphal    Håkan Grahn    Niklas Lavesson

Blekinge Institute of Technology  
Karlskrona, Sweden

flw@bth.se

2016-02-02

# Outline

- 1 Document Readability
- 2 Howe's Binarization Algorithm
- 3 Heterogenous Computing
- 4 Binarization Pipeline

- Swedish university, established in 1989
- Over 6000 registered students
- BigData@BTH
- Swedish company, established in 2004
- Provides access to almost 60 million images
- Church books, court records, military records, census records, ...



in real life



# Document Readability

Ref. No. 1136 N° 19		Ref. No. 1136 N° 19	
4.2		J. Bengtsson	500 11.5:6-24473
4.2	John & wife	G. Hammar	7.41531221
4.2	John & wife	G. Larsson	7.6.57552442
4.2	John & wife	J. Larsson	5.5523
4.2	John & wife	J. Pernilsson	4.741185.2
4.2	friends	J. Nilsson	7.851744242
4.2	(N. 44)	G. Elma	5.1744442
4.2	Wane	J. Nils	7.11.2442
4.2	W. E. Karpas	J. Persson	8.
4.2		O. Olaf	1.
4.2		J. Larsson	1/2
4.2		G. J. G. - O. Berg	4.841744242
4.2		Peter Nilsson	40.1 48.644442
4.2	N. 18	P. Bengtsson	48.62644442
4.2	John & wife		

# Approach

Now if there be any meaning in words, all  
as a legislator prescribing or a profet for explaining  
this must signify, that the moral conduct of a  
Laws is just a Being as this. -

# Approach

Now if there be any meaning in words, all  
as a legislator prescribing or a profet for explaining  
this must signify, that the moral conduct of a  
Laws to such a Being as this. -

Now if there be any meaning in words, all  
as a legislator prescribing or a profet for explaining  
this must signify, that the moral conduct of a  
Laws to such a Being as this. -

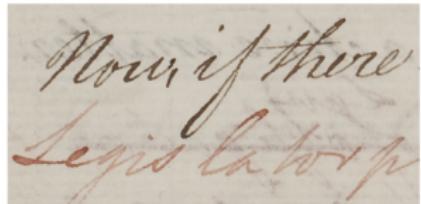
# Approach

Now if there be any meaning in words, all  
as a legislator prescribing or a profet for explaining  
this must signify, that the moral conduct of a  
Laws to such a Being as this. -

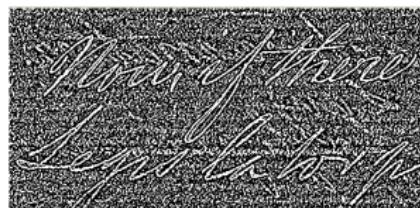
Now if there be any meaning in words, all  
as a legislator prescribing or a profet for explaining  
this must signify, that the moral conduct of a  
Laws to such a Being as this. -

# Approach - Demo

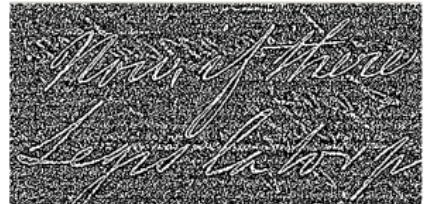
# Howe's Binarization Algorithm



Now if there  
Legis lator p

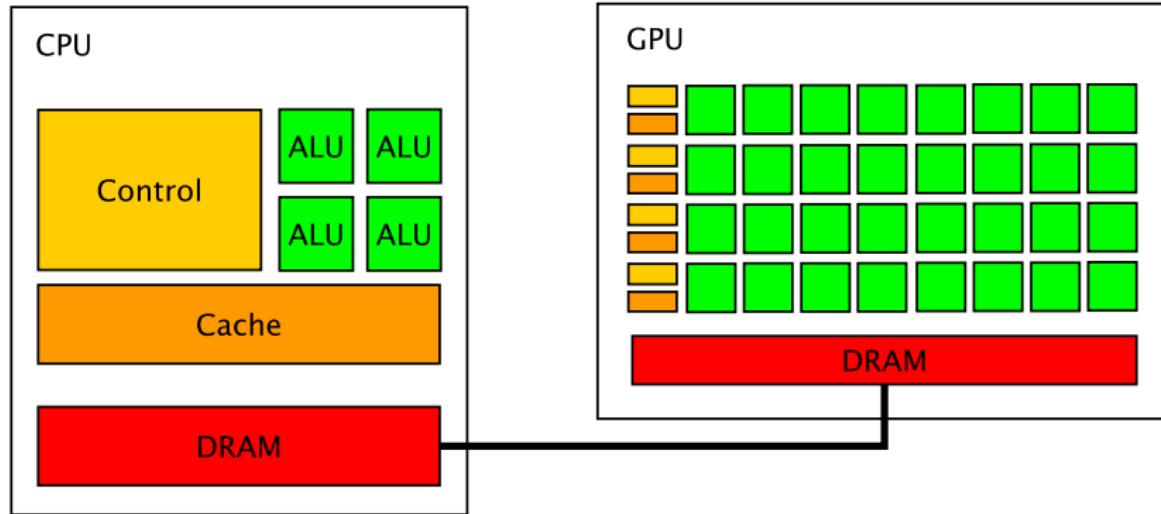


# Howe's Binarization Algorithm (Cont.)

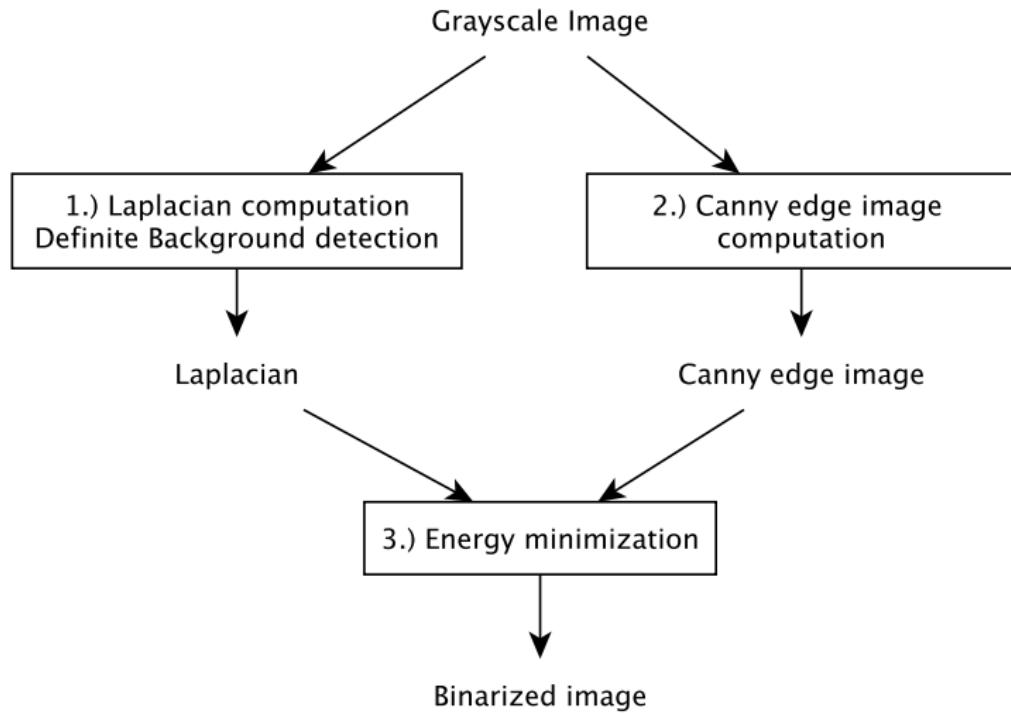


Now, if there  
Leads to work

# Heterogenous Computing



# Binarization Pipeline



# Binarization Pipeline (Cont.)

	I	II	III	IV	V	VI	VII	VIII
1	CPU	GPU	CPU	GPU	CPU	GPU	CPU	GPU
2	CPU	CPU	GPU	GPU	CPU	CPU	GPU	GPU
3	CPU	CPU	CPU	CPU	GPU	GPU	GPU	GPU

# Preliminary Results

## Reference Implementation

Now if there be any meaning in words, all  
as a legislator prescribing or a profet for explaining,  
this must signify, that the moral conduct of a  
laws is fit with a Being as this. -

## Configuration IV

Now if there be any meaning in words, all  
as a legislator prescribing or a profet for explaining,  
this must signify, that the moral conduct of a  
laws is fit with a Being as this. -

# Preliminary Results (Cont.)

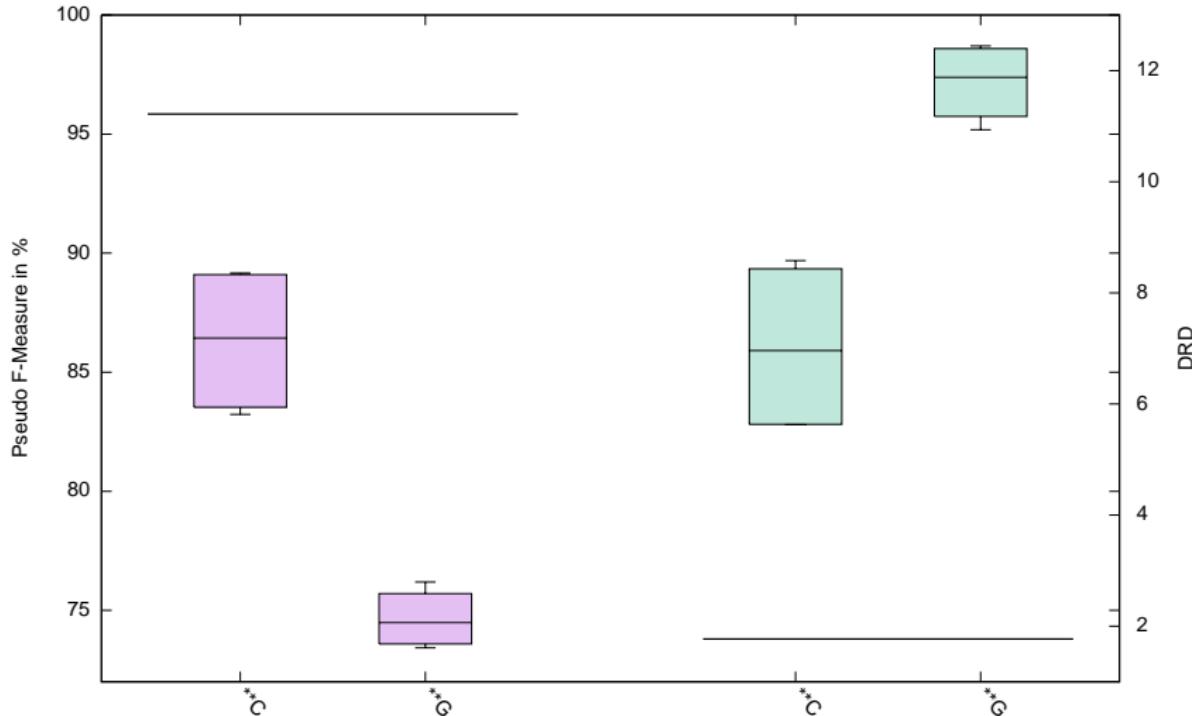
## Reference Implementation

Now if there be any meaning in words, all  
as a legislator prescribing or a profet for explaining  
this must signify, that the moral conduct of a  
law is with a Being as this. -

## Configuration VIII

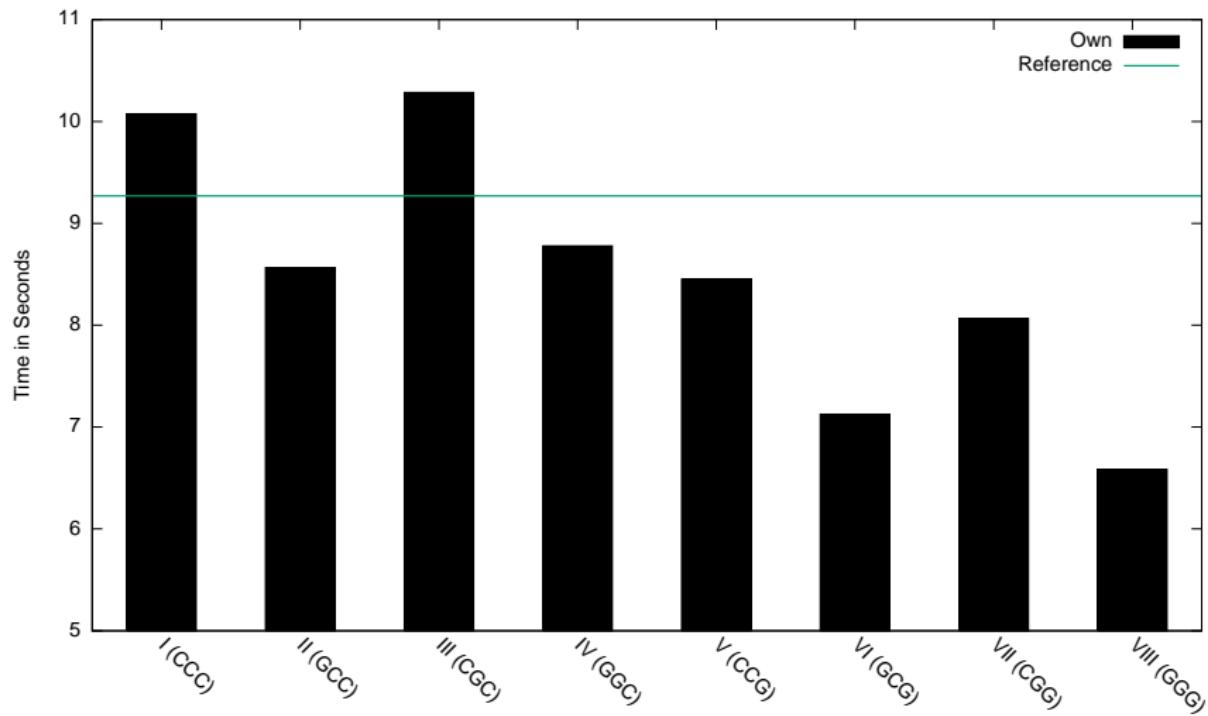
Now if there be any meaning in words, all  
as a legislator prescribing or a profet for explaining  
this must signify, that the moral conduct of a  
law is with a Being as this. -

# Preliminary Results - Binarization Performance



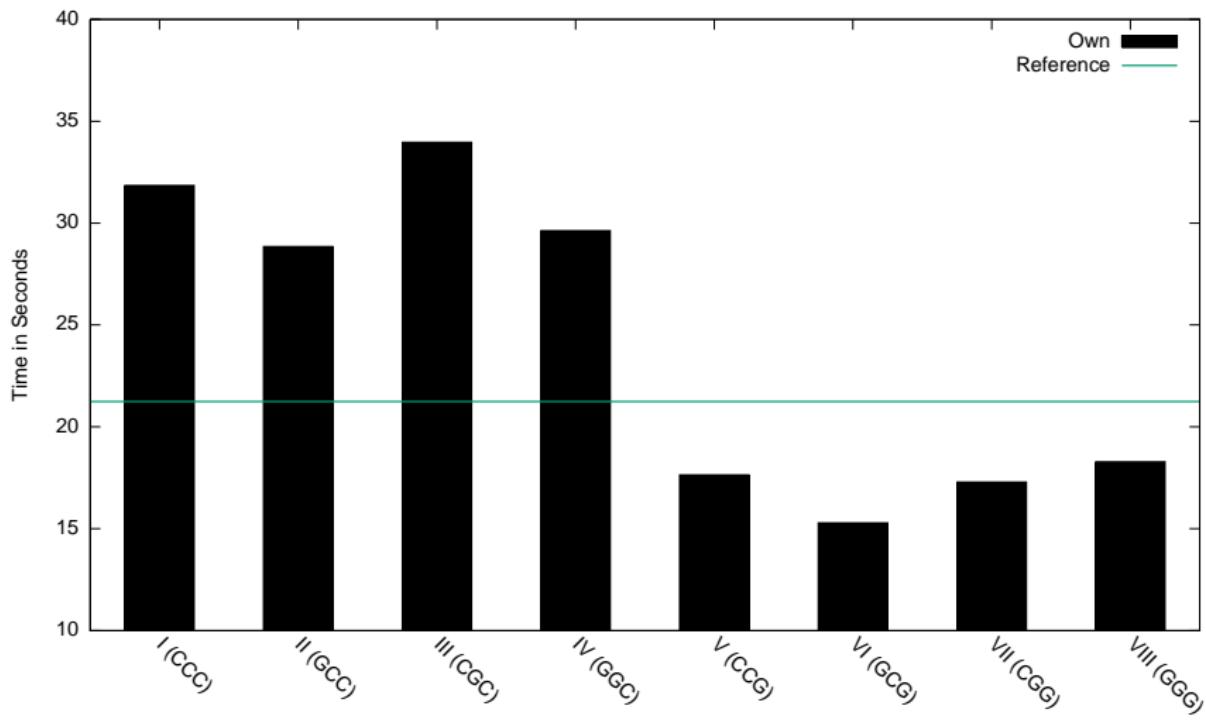
# Preliminary Results - Time

H-DIBCO 2014 Benchmark



## Preliminary Results - Time (cont.)

High Resolution Image



# Preliminary Results - Time per Step

Time taken in each binarization step  
for the used high resolution image.

	1	2	3
CPU	2.27 s	0.17 s	28.76 s
GPU	0.39 s	0.11 s	14.54 s

# Next Steps

- Revision of the implementation
- Implementation of the binarization pipeline

# Acknowledgements

We would like to thank ArkivDigital for providing us with access to their image database.

This work is part of the research project "Scalable resource-efficient systems for big data analytics" funded by the Knowledge Foundation (grant: 20140032) in Sweden.