GmlmgProc

Alexandra OlteanuSCPDAlexandru ŞtefănescuSCPD

Image processing for OCR and content extraction

Implies a set of costly operations like:

- Binarization – the resulted image is formed by two classes separated by a threshold *t* : foreground and background

- Noise detection
- Clustering for creating meta-entities (ex: characters, lines)
- Detection and/or classification of page components: line, table,

character, frame, image, etc

- Skew Detection and correction
- Geometrical distortions correction
- And the list may continue

For this project our aim was to implement methods for image binarization and to try different solutions to improve our results and to remove noise:

- Otsu 1D
- Otsu 2D
- Niblack
- Color Uniformization

Input (grayscale)

gust 28, 1905, in Brooklyn.

"The witnesses were William Moore and a Mr. Toy. In September, a few days previous to her brother's marriage, Mrs. Carlton had some words with her parents, and she and her brother left home, the brother occupying my room at 198 Sands street, until his marriage. After one week at the Hotel St. George Mrs. Carlton went to board in East New York. She remained there until November, when she joined me in Washington, She visited Brooklyn in February, remaining one month.

"In May Mrs. Carlton, in putting on a felt slipper with a rubber sole, stamped on the floor, and a needle sticking in the matting ran through the sole into her foot and broke off. It was extracted. Two days later she went with me to the Zoological Gardens. Her foot paining her, she removed her shoe and stocking and walked barefooted for two

GENERAL BLACKMAR

Assumes Headship of the Grand Army of Republic.

WILL ISSUE GENERAL ORDER

New Commander-in-Chief Expresses Great Grief at Death of Predecessor.

Otsu 1D

rengéed we write sach historieles ét. Rostin, bride Briekiyn

TTP: Witnesse zere Willin: Melte and a Millin in September, offer date provide to a boother structure Mrs. South to a boother words with her is anyther and showed ber another 1970 conclusted with the supplying roomer 1970 to set to brother securitying roomer with the South structure Hard Structure Mrs. Coller wert to board in Hast New York. She tempined there at 2 Notestboth when wert to board in Hast New York. She tempined there at 2 Noteststructure parted me in Workhouston, She bished Brook, which Hast Denter and another end month.

(iii) M. S. Miss, Carltons in patting on a test support with a public wede, at opport of the fluor, and is modified to a give the upport garage through the sole at an feet and brake off. It was extended.

(a) Two days later size went with the term is Zonogenel Gardens. Here for a prining for She remoted her show to in Socklag and Wallow brocketed to two

GENERAL BLACKMAR

Assumes Headship of the Grand Army of Republic.

WILL ISSUE GENERAL ORDER

New Commander-in-Chief Expresses Great Grief at Death of Predecessor.

.....

Otsu 2D

recorded we were again married on A+1 gust 28, 200, in Brooklyn.

"The witnesses were William Moore and a Mr Toy. In September, a few days previous to her brother's marriage. Mrs. Curiton had some words with her barents, and she and her brother left some, the brother occupying my room at 198 Sands street, until his marriage. After one week at the Hotel St. George Mrs. Carlton went to board in East New York. She remained there until November, when she joined me in Washington, She visited Brooklyn in February, remaining one month.

"In May Mrs. Carlton, in putting on a felt slipper with a rubber sole, stamped on the floor, and a needle sticking in [the matting ran through the sole into beinfoot and broke off. If was extractied. Two days later she went with me to the Zoological Gardens. Her foot paining her, she removed her show and storking and walked barefooted for two hears. Referring home foot.

GENERAL BLACKMAR

Assumes Headship of the Grand Army of Republic.

WILL ISSUE GENERAL ORDER

New Commander-in-Chief Expresses Great Grief at Death of Predecessor. Niblack

recorded we were again married on A gust 23, 1977, in Brooklyn, 2022 The i witnesses | were William Moore and a Mr. Toy, 7 In September, a few days previous to her brother's marriage, Mrs. Carlton had some words with her parents, and she and her brother left home, the brother occupying my room at 198 Sands street, until his marriage. After one week at the Hotel SLE George Mrs.' Carlton went to board in East New York. She remained there until November, when she joined me in Washington, She visited Brooklyn in February, remaining one month. Have a reason of the ""In May Mrs. Cariton, in putting on a felt slipper with a rubber sole, stamped on the floor, and a needle sticking in the matting ran through the sole into her foot and broke off. It was extracted. Two days later she went with me tes the « Zoological - Gardens, ». Her (foot | naining her. she removed her stocking and mathed benefact.



.....

Color Uniformization (intermediate grayscale)

recorded we were again married on Au gust 28, 1903, in Brooklyn.

"The witnesses were William Moore and a Mr. Toy. In September, a few days previous to her brother's marriage. Mrs. Carlton had some words with her parents, and she and her brother left home, the brother occupying my room at 198 Sands street, until his marriage. After one week at the Hotel St. George Mrs. Carlton went to board in East New York. She remained there until November, when she joined me in Washington. She visited Brooklyn in February, remaining one month.

"In May Mrs. Carlton, in putting on a felt slipper with a rubber sole, stamped on the floor, and a needle sticking in the matting ran through the sole into her foot and broke off. It was extracted. Two days later she went with me to the Zoological Gardens. Her foot paining her, she removed her shoe and stocking and walked berefected for two

N	-	and in the local division of the local divis	Lun	n.	-	Norid	1	antest of	-	A	V			HALL I
1	T.	CCC.	A and	The second	a state		5	Contraction of the local division of the loc		NCHINA PARA	R.	Number of Street, or other	The state	
UA.	5	1.1	S	1 11	1	-	0	5	3	U	11	1	11	

Assumes Headship of the Grand Army of Republic.

WILL ISSUE GENERAL ORDER

w Commander-in-Chief Expresses Great Grief at Death of Predecessor.

Color Uniformization + Otsu 1D

recorded we were again married on $A \neq \frac{1}{4}$ gust 23, 1993, in Brooklyn.

"The witnesses were William Moore and a Mr. Toy. In September, a few days previous to her brother's marriage. Mrs. Carltop had some words with her parents, and she and her brother left home, the brother occupying my room at 198 Sands street, until his marriage. After one week at the Hotel St. George Mrs. Carlton went to board in East New York. She remained there until November, when she joined me in Washington. She visited Brooklyn in February, remaining one month.

"In May Mrs. Carlton, in putting on a felt slipper with a rubber sole, stamped on the floor, and a needie sticking in the matting ran through the sole into her feot and broke off. It was extracted. Two days later she went with me to the Zoological Gardans. Her foot paining her, she removed her shoe and stocking and walked harefooted for two



Assumes Headship of the Grand Army of Republic.

WILL ISSUE GENERAL ORDER

Tew Commander-in-Chief Expresses Great Grief at Death of Predecessor.

Otsu 1D

Otsu method is based on a very simple idea: find the threshold that minimizes the weighted within-class variance, defined as a weighted sum of variances of the two classes.

$$\sigma_w^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t)$$

- ω_i are the probabilities of the two classes separated by a threshold *t*
- σ_i^2 are variances of these classes

This can be demonstrated to be the same as maximizing the between-class variance.

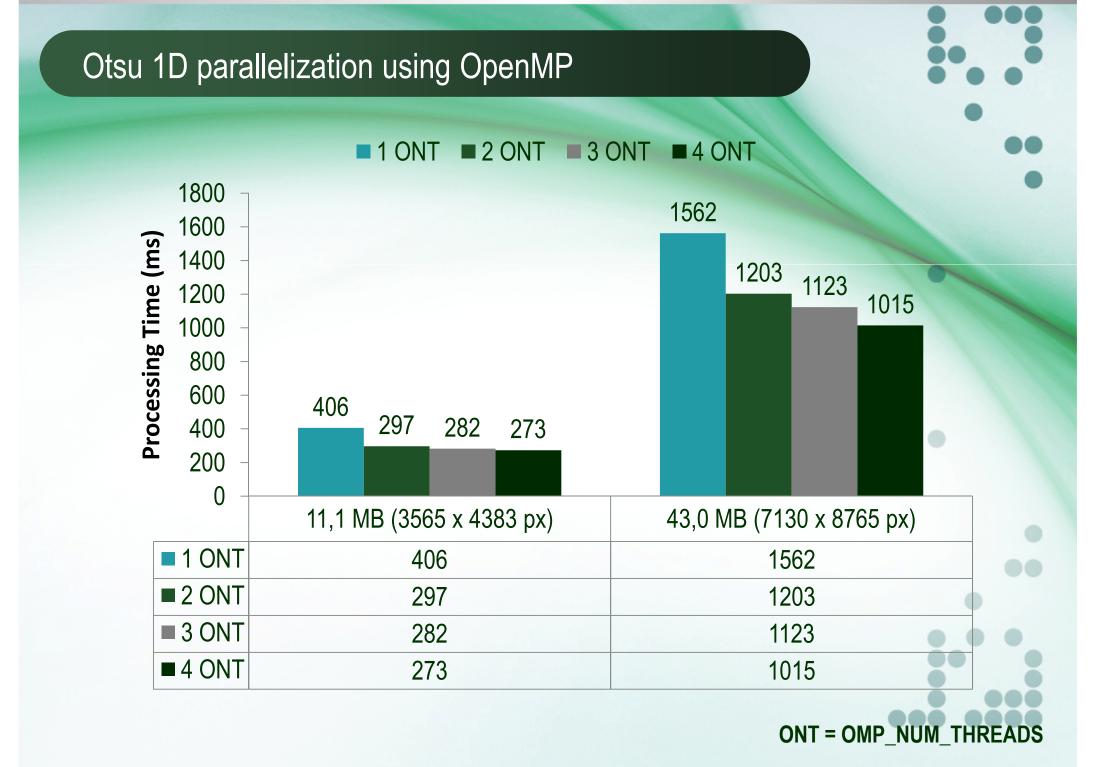
$$\sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) = \omega_1(t)\omega_2(t) \left[\mu_1(t) - \mu_2(t)\right]^2$$

where μ_i represent the class means

Implementation: we simply compute the class means and their probabilities

Advantages:

This method is fast because operates directly on the gray level histogram. It provides very good results when the numbers of pixels in each class are close to each other.



✓ granularity => good \checkmark fast serial algorithm => low speedup ✓ load balance => good

Profiling results:

✓ CPU time utilization => very good \checkmark hot spot => histogram computation

	m	Elapsed Time: CPU Time: Wait Time: Wait Count: Unused CPU Time Core Count: Threads Created	2	,	••
			~		•
Elansed Time		Average CPU Usag	Target Concurrenc		
	Ic	0 i Poor Corrections Runn	2 3+ deal Ove		
>	of about five n of introduction of introduction of the second of the second of the second days previous to Mrs. Carlton be mome. the broth mome. the broth mome. the broth mome. the broth mome. the broth mome. The broth mome. The second of the second of the main and the second between the second mainting one mome for the second of the se	o her brother's marriage, ad some words with her he and her brother left er occupying my room at , until his marriage. Af- the Hotel St. George the to board in East New ained there until Notem- oined me in Washington.	ING SUCCE GENERAL I ssumes Head Frand Army o ILL ISSUE GEN	EDS BLACKMAR Iship of the of Republic.	••••

Commander-in-Chief Expresses

Great Grief at Death of

Predecessor.

ed we were again married on A+1 st 28, 1927, in Brooklyn. The willnesses were William Moore id a Mr. Toy. In September, a few ys previous to her brother's marriage, top had some words with he nts, and she and her brother lef other occupying my room at and arrenter wentyping my room in one week at the Hoad S. George Carlton went to hoard in Easy New Men seb joined mer unit Novem-when seb joined mer unit Novem-traited Exostopin in Pobrary, re-WILL ISSUE GENERAL ORDER

ke off.



Assumes Headship of the

WILL ISSUE GENERAL ORDER

New Commander-in-Chief Expresses Great Grief at Death of Predecessor.





Assumes Headship of the Grand Army of Republic.

WILL ISSUE GENERAL ORDER

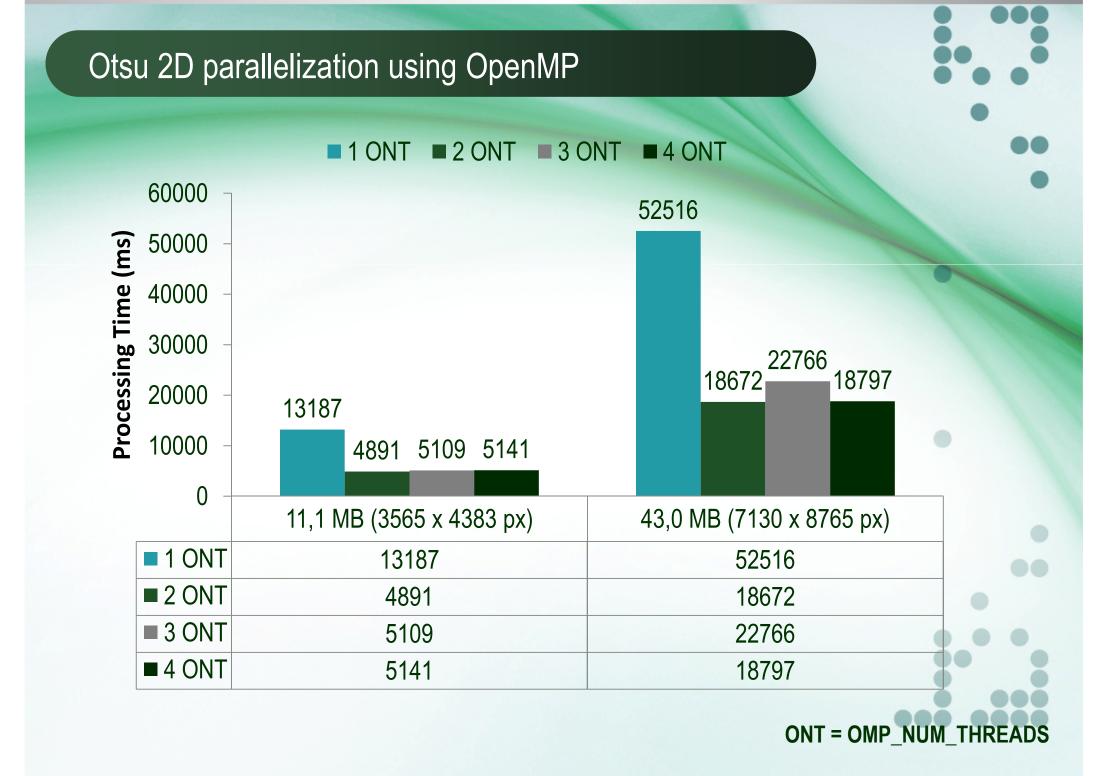
The particular indication examples in the second se New Commander-in-Chief Expresses later sn.

- >considers both the gray value of a pixel and the average gray value of it neighborhood
- >pixels are partitioned into two classes by a threshold pair (s, t), where s is gray level of
- a pixel and t it's local average gray level
- >Otsu 2D method is based on the same principles as classic Otsu
- ➢it uses the two dimensional histogram
- ➢ by using the trace of the between-class variance matrix as the measurement of between
- class variance and maximizing it we select the corresponding threshold (s, t) for this value.

Implementation: we use a fast recursive algorithm to implement Otsu 2D method.

Advantages:

- more robust to noise than Otsu 1D
- provides very good results when the numbers of pixels in each class are close to eachother



✓ in OpenMP 2.5 and in earlier versions the nested parallelism can be exploit only by using nested parallel regions

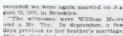
✓ granularity => good

✓ atomic constructs introduce synchronization overhead

 \checkmark load imbalance – conditional instructions

Profiling results:

 \checkmark hot spot => local average gray level computation





Assumes Headship of the Grand Army of Republic.

WILL ISSUE GENERAL ORDER

Commander-in-Chief Expres Grief at Death of





Assumes Headship of the Grand Army of Republic.

WILL ISSUE GENERAL ORDER Lew Commander-in-Chief Expresses Great Grief at Death of Predecessor.



Elapsed Time: 11.843s CPU Time: 17.664s Wait Time: 10.850s Wait Count: 28012 Unused CPU Time: 1.810s Core Count: 2 Threads Created: 20 1.85 CPU Usage

Simultaneous Running Threads

Idea



Assumes Headship of the Grand Army of Republic.

3+

Over

WILL ISSUE GENERAL ORDER

Great Grief at Deat Predecessor.

Otsu 2D tuned with OpenCL



CPU: Intel Core 2 Duo, T7500, 2. 2 GHz, 1.99 GB RAM - Display Mobile Intel(R) 965

	Ol	penCL (secon	ds)	Serial (seconds)			
Image size	11 MB	44 MB	62 MB	11 MB	44 MB	62 MB	
Transform data into CL structures	0.438	1.735	1.765				
Create CL context, buffers etc.	0.5	0.421	0.453				
Runtime	0.953	3.094	3.141	2.937	11.672	11.859	
Copy computed data from CL structures into app structures	0.156	0.641	0.719				
TOTAL time	2.781	9.359	9.547	3.344	13.297	13.453	

Helped by Iulian Milas 🙂



Niblack's method is a local threshold binarization method which involves computing for each pixel in the grayscale image the mean and standard deviation of the colors of the neighboring pixels in an area (window) of predefined size. The threshold for determining if the pixel will be black or white is computed by the formula:

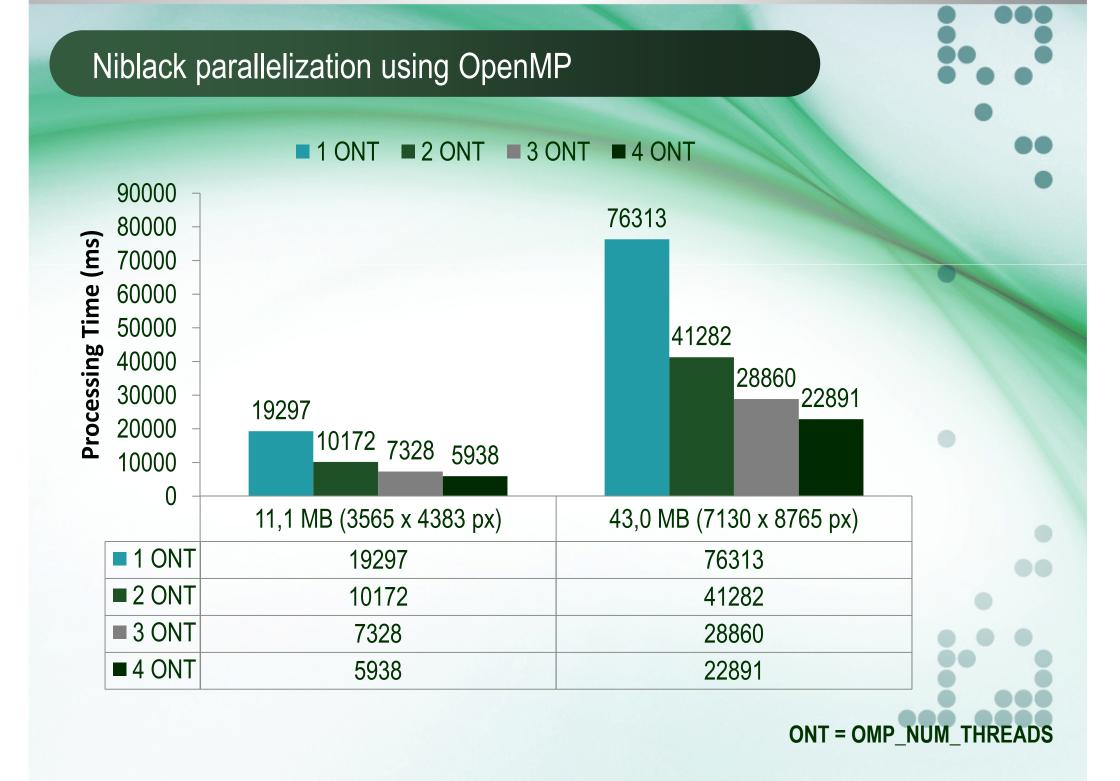
T = mean + k * stdev

where k is a constant, preselected coefficient.

Implementation: We simply slide the window onto each pixel and recompute the mean and standard deviation.

Assessment:

- does not involve a unique threshold for the entire image and thus is adaptable for local particularities
- had worst results in our tests, despite being considered the best binarization method for text images
- is much slower than other methods.



- ✓ granularity => good
- \checkmark trivial to parallelize
- \checkmark each iteration is completely independent on the rest
- ✓ load balancing easy to achieve
 ✓ CPU wait time practically zero (except for I/O)

	Elapsed Time: CPU Time: Wait Time: Wait Count: Unused CPU Time: Core Count: Threads Created: 1.87	16.381s 29.381s 0.918s 289 2.195s 2 2 2	••
	ĺ		•
Elapsed Time	Average CPU Usage	Target Concurrency	
	0 1 2 Idle Poor Ide Simultaneous Runnin	al 🚺 Ove	
	onnartan cous rtani ini	ig micaus	

recorded we were equal married on Asrout 5, 197, in Brooklyn. "The witnesses were William Mooreand a Mr. Toy. In September, a few days previous to her brother's marriage. Mrs. Cariton had some words with her sarents, and she and her brother left more, the brother occupying my room at home. the brother occupying my room at fer one week at the Hotel St. George Mrs. Cariton went to board in East New

York. She remained there until November, when she joined me in Washington She visited Brooklyn in February, remaining one month. "In May Mrs. Cariton, in putting on a

felt slipper with a rubber sole of sumped on the floor, and a media sticking in the matting rand a media sticking in the matting rand broke for the sticking of the foot and broke of the sticking sticking of the sticking and broke sticking sticking and the sticking and walked broadcated for the stocking and walked broadcated for the



Assumes Headship of the Grand Army of Republic.

WILL ISSUE GENERAL ORDER

r Commander-in-Chief Expresses Great Grief at Death of Predecessor.



recorded we were again married on AA gut 23, 1971, in Brookiyn, were a series of the control of the series of the series of the series of the days previous to her brothers amarting, days previous to her brother series of the parents, and she and her; brother, left marring, and she and her; brother, left and a she series, until his mariage, Af parents, and she and her; brother, left and a she series of the series of the parents, and she and her; brother, left and a she series of the series of the parents, and she series of the parents, and she series of the series of the series of the series of the parents, and she series of the brother of the series of the floor, and a needle sticking in the marking frag through the series the series of the series

General BLACKMAR Harrison at Geore Geore Grand Army of Republic WILL'ISSUE GENERAL'ORDER WILL'ISSUE GENERAL'ORDER WILL'ISSUE GENERAL'ORDER Mer 2 Commander-in Chief 2 Expressed

and housi the sole into New Commander-in-Chief Expresses Tan through the sole into New Commander-in-Chief Expresses ya later also went with meoptenic Garaten as: Her for the Great Grief at Death of Mer are moved her shoe and tangent the sole of the so



Image color uniformization

- is a preprocessing technique for grayscale images before the binarization step.
 creates an intermediate grayscale image which binarized with global methods like
 Otsu 1D or Otsu 2D produces better results than without the preprocessing
 the main idea is to "diffuse" color differences weighted by distances between
- the main idea is to "diffuse" color differences weighted by distances between neighboring pixels
- each pixel in the intermediate image is computed using the following sum, for the pixels in a neighboring area (window) of predefined size:

$$\sigma(x, y) = \sum_{(x', y') \in W} \frac{color(x', y') - color(x, y)}{\sqrt{(x - x')^2 + (y - y')^2}}$$

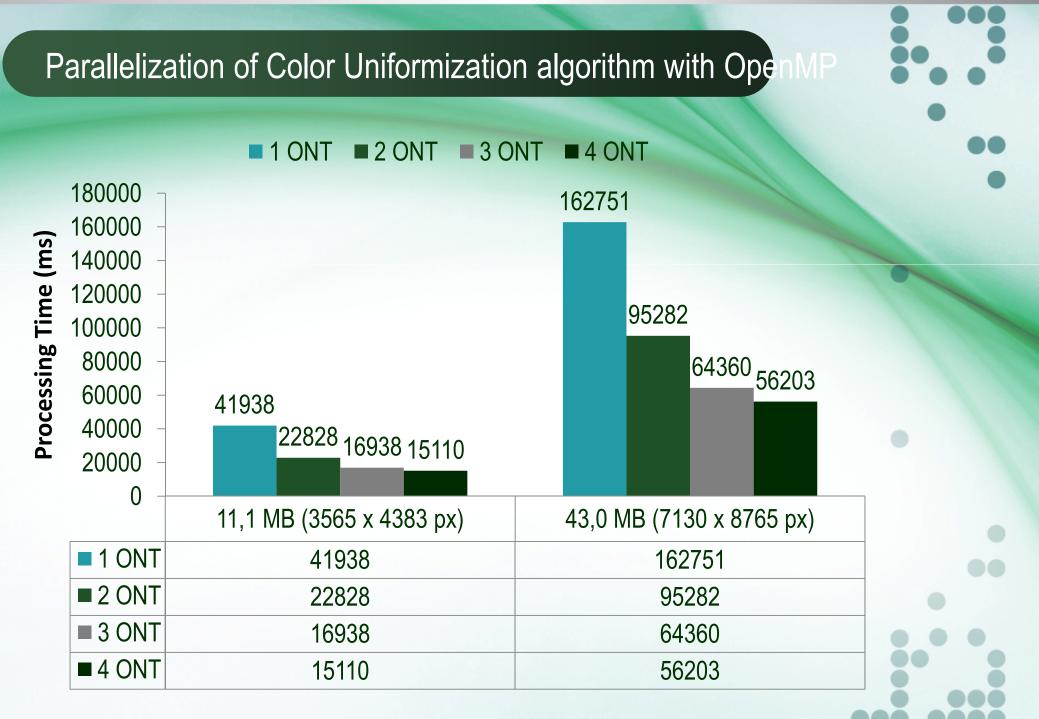
for every pixel of coordinated (x', y') located in the window *W* centered in (x, y)in the end, the sums are scaled to 0 - 255 to correspond to a level of gray.

Implementation: We simply slide the window onto each pixel and recompute the mean and standard deviation.

Assessment:

 obtained a great improvement in the image binarized with Otsu 1D and Otsu 2D using this preprocessing step as compared to binarizing without it.

• similar to Niblack's method, it is very slow.



ONT = OMP_NUM_THREADS

- similar to Niblack's method
- \checkmark granularity => good
- ✓ each iteration is completely independent on the rest
- ✓ CPU wait time practically zero (except for I/O)

Profiling results: \checkmark hot spot => computing the sum

led we were again married on A+ zust 28, 1903, in Brooklyn, The witnesses were William Moore d a Mr. Toy. In September, a few vious to her brother's marriage had some words with he nd she and her brother lef brother occupying my room at Assumes Headship of the he week at the Hotel St. George Grand Army of Republic. went to board in East New Mrs. Carlton, in putting on a with a rubber sole, stamped and a needle sticking



Commander-in-Chief Expresses Great Grief at Death of

Predecessor







Mrs. Carlton had some words with her arrents, and so and her bother left nome, the boother occupying my room at BS Stude streat, until his muriling. Af-try one work at the Hotel St. General Ort. She primined these unil Notem-ber wither Stooking in Footmary, re-"In May Mrs. Cariton, in putting on a Cit slipper with a rubbic sole, stimped on the floor, and a medie sticking in the floor state abe went with me the floor state abe went went went went went we the floor state abe went we the floor state ab

	Elapsed Time: CPU Time: Wait Time: Wait Count: Unused CPU Time: Core Count: Threads Created:	37.473s 66.090s 2.484s 636 3.636s 2 2 2	
Elapsed Time	Average CPU Usage	Target Concurrency	
	0 1 2 dle Poor Ide Simultaneous Runnin	al Ove	_
resorded we were agg gust 23, 10%, in Brook "The witnesses we and a Mr. Toy. In days previous to ber		BLACKMA	

Conclusions (what we've learned!)

OpenMP:

- reducing false sharing => making use of private data as much as possible
- parallelize at the highest level possible
- > make parallel regions as large as possible to reduce parallelization overhead
- use CRITICAL sections for fine-grained locking
- > the work should be distributed evenly across the threads
- > **NOWAIT** clause => to eliminate redundant or unnecessary barriers
- ➤ we also compiled with Intel C++ but we obtained similar results (but we hated it)
- ➢ profiling was made on a two core system

OpenCL:

- ➤ very fast
- > does not support bidimensional arrays
- > setting up the OpenCL kernel for running is very complicated



..and autographs