

Boosting OCR Classifier by Optimal Edge Noise Filtering

***Junior Barrera
Marcel Brun
Routo Terada***

Universidade de Sao Paulo,
Departamento de Ciencia da Computação

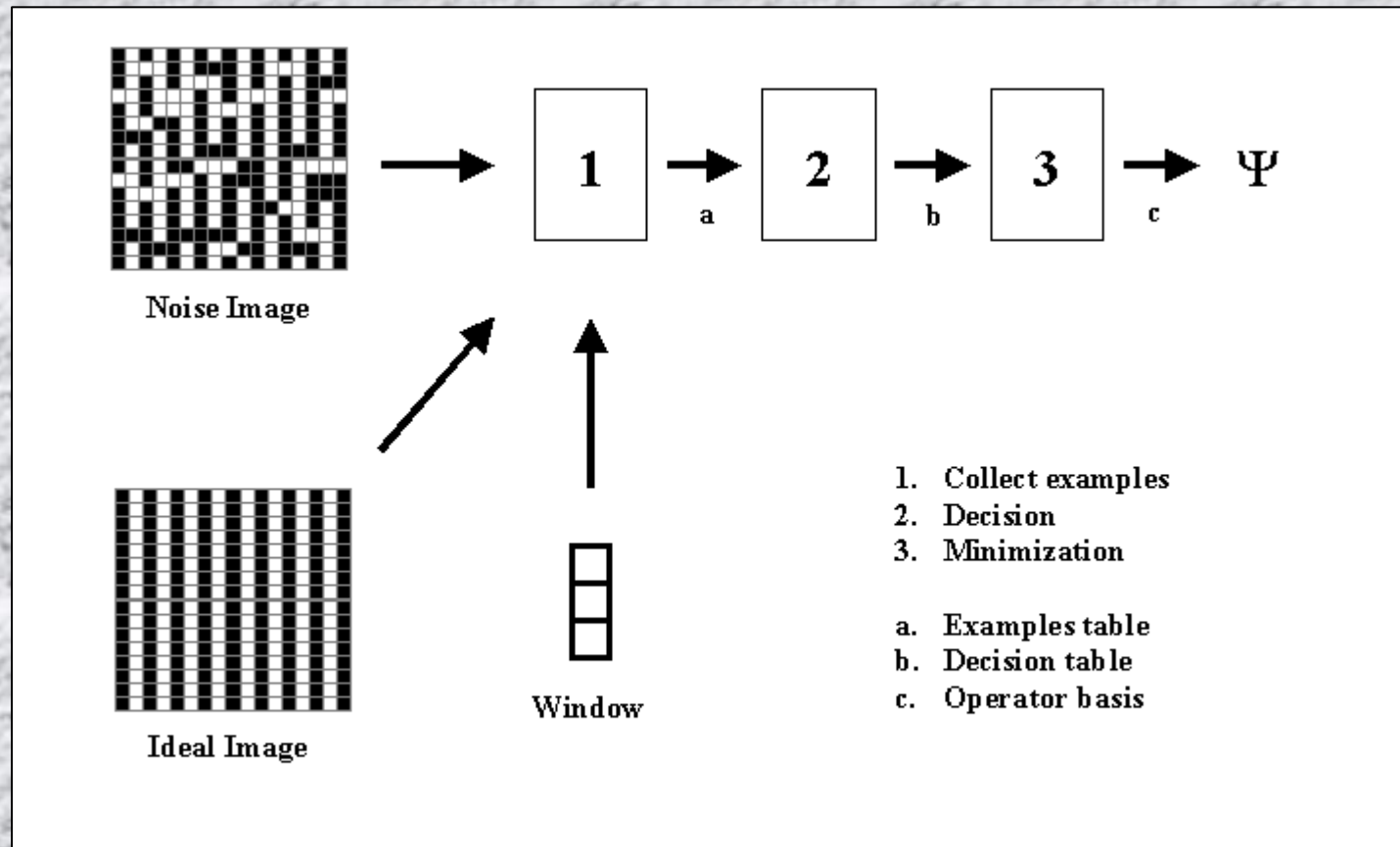
Edward Dougherty

Texas A&M University,
Department of Electrical Engineering

Outline

- Operator Design
- Character Normalization
- Problems in resolution reduction
- Anchoring
- Edge Noise Filtering
- Experimental Results
- Conclusions

Operator Training



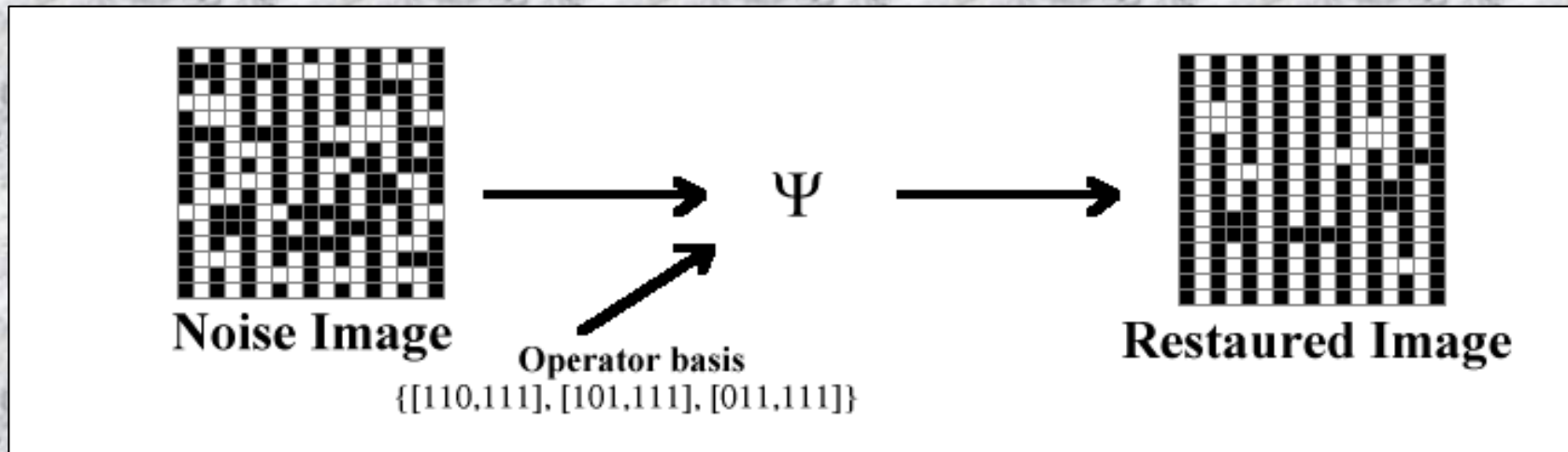
Complete scheme

Operator Training

$x_1x_2x_3$	Frequency of 0	Frequency of 1	$x_1x_2x_3$	$h(x)$
0 0 0	86	0	0 0 0	0
0 0 1	19	2	0 0 1	0
0 1 0	18	0	0 1 0	0
0 1 1	1	16	0 1 1	1
1 0 0	19	2	1 0 0	0
1 0 1	0	14	1 0 1	1
1 1 0	1	16	1 1 0	1
1 1 1	0	78	1 1 1	1
	a			b

Collect and decision

Operator Training



Application

Classification steps

quantos esta publica escritura virem que, no ano da Era Cristã, de mil novecentos e noventa e oito (1.998), ao **DEZESETE** (17) dia do mês de **NOVEMBRO**, nesta cidade de São Paulo, em meu Cartório, perante mim, Tabelião, compareceram partes entre si, justas e contratadas, outorgantes e

Binary Image

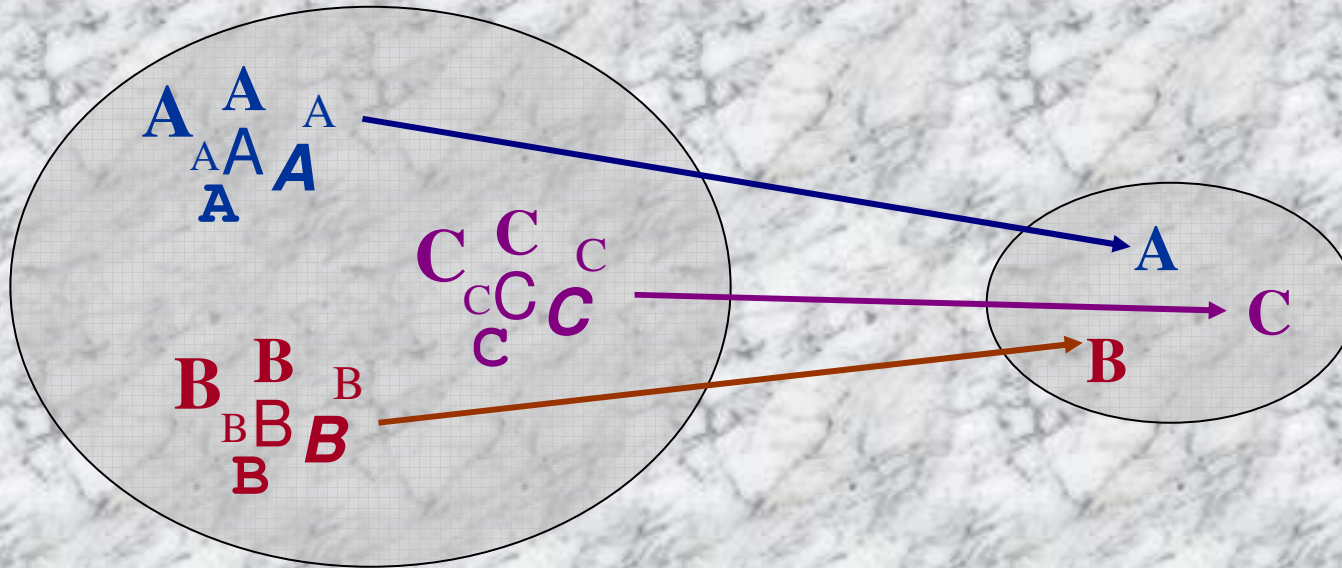
quantos esta publica escritura virem que, no ano da Era Cristã, de mil novecentos e noventa e oito (1.998), ao **DEZESETE** (17) dia do mês de **NOVEMBRO**, nesta cidade de São Paulo, em meu Cartório, perante mim, Tabelião, compareceram partes entre si, justas e contratadas, outorgantes e

Labeled Image

quantos esta publica escritura virem que, no ano da Era Cristã, de mil novecentos e noventa e oito (1.998), ao **DEZESETE** (17) dia do mês de **NOVEMBRO**, nesta cidade de São Paulo, em meu Cartório, perante mim, Tabelião, compareceram partes entre si, justas e contratadas, outorgantes e

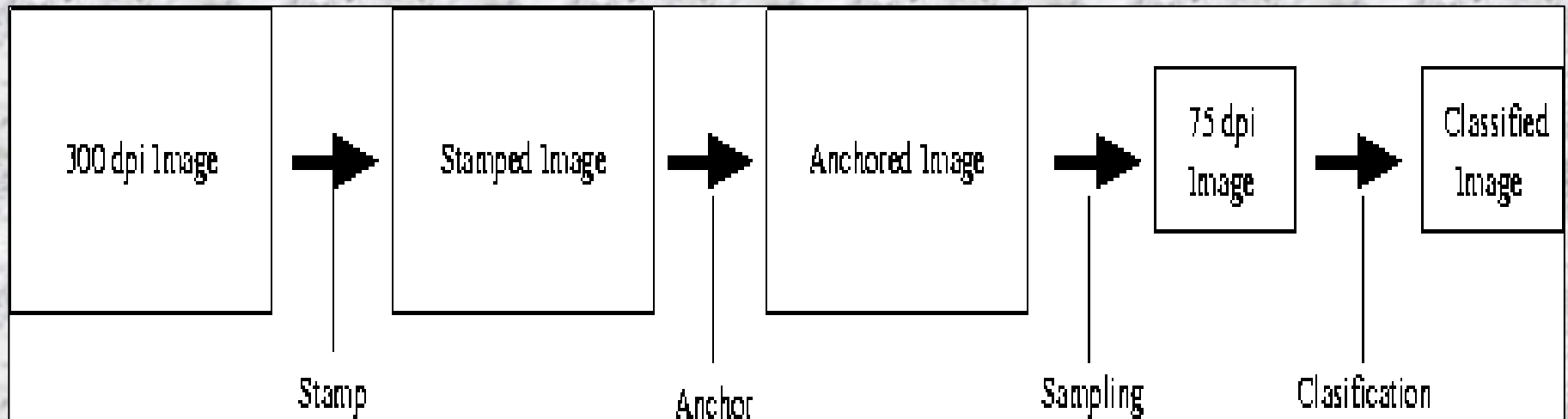
Classified Image

Character Normalization



Normalizing characters, the work of differentiate them becomes easier

Character Normalization

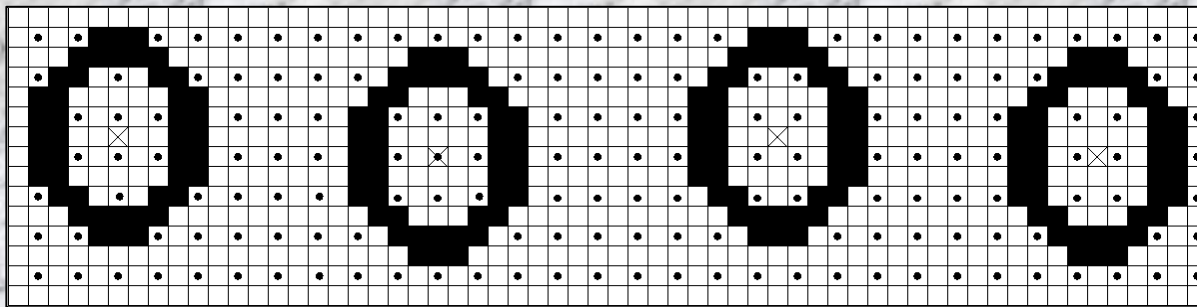


Normalisation Diagram

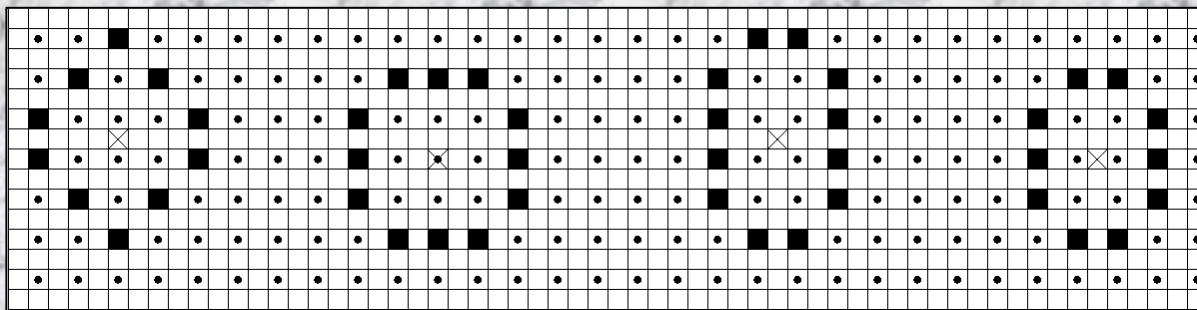
- ◆ The complexity grows with the windows W
- ◆ Anchoring \Rightarrow Decrease image resolution
- ◆ Stamp \Rightarrow Normalise characters shape

Problems in resolution reduction

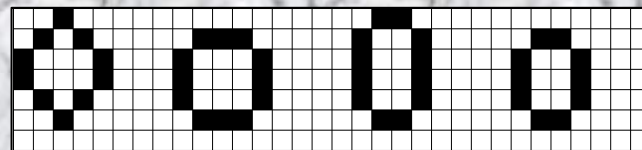
- **Objets with same shape, at different positions may turn into different shapes at low resolution.**



Original Image



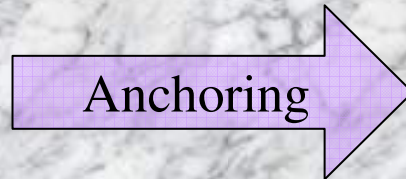
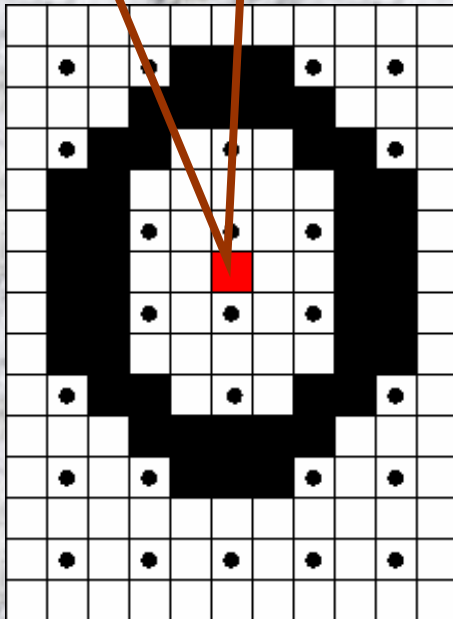
Sampling



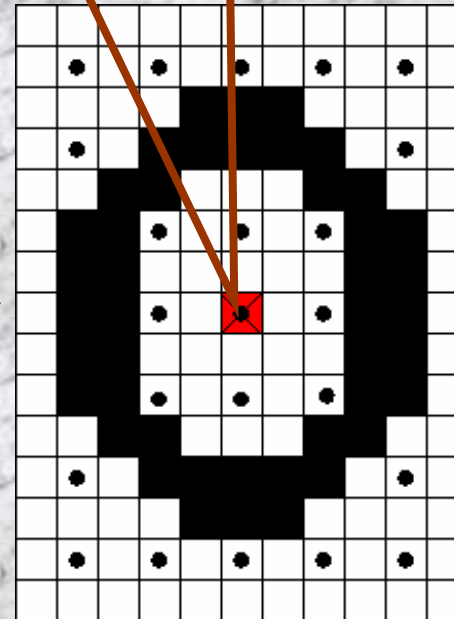
Shrinking

Anchoring

Mass center (or anchor)
outside the sampling grid

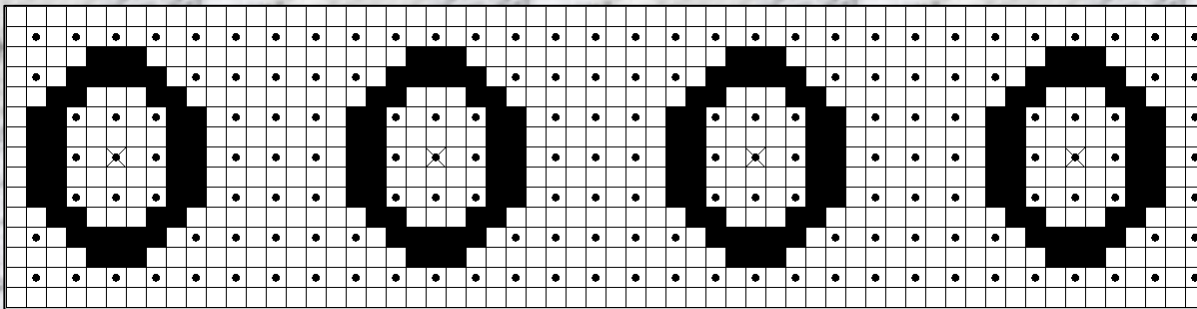


Mass center (or anchor)
inside the sampling grid

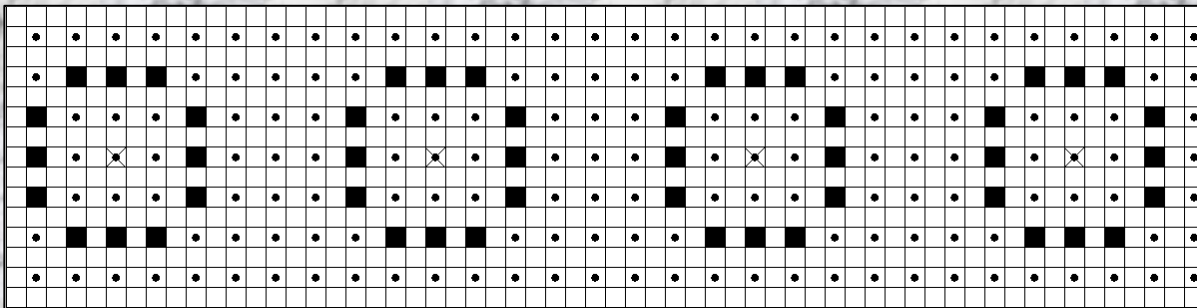


Anchoring

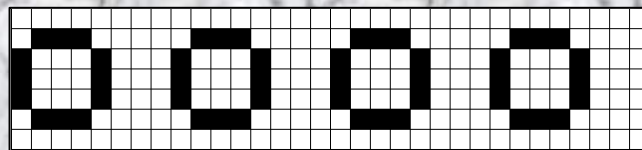
- It consist of traslating the mass center of each object over the low resolution grid



Original Anchored Image



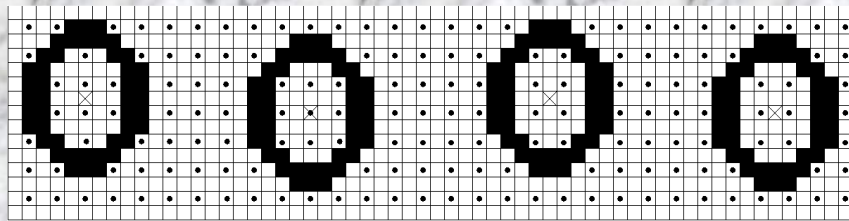
Sampling



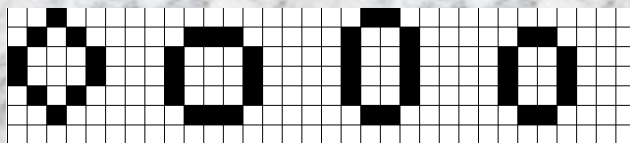
Shrinking

Anchoring - Comparison

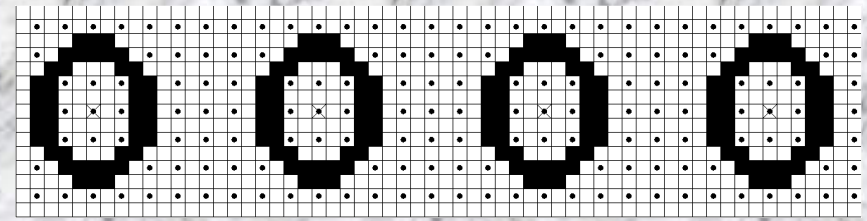
Without Anchor



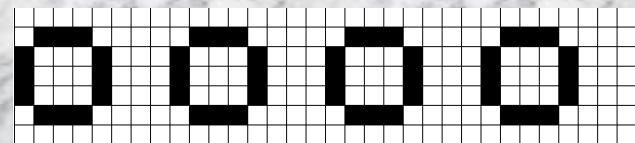
Shrink



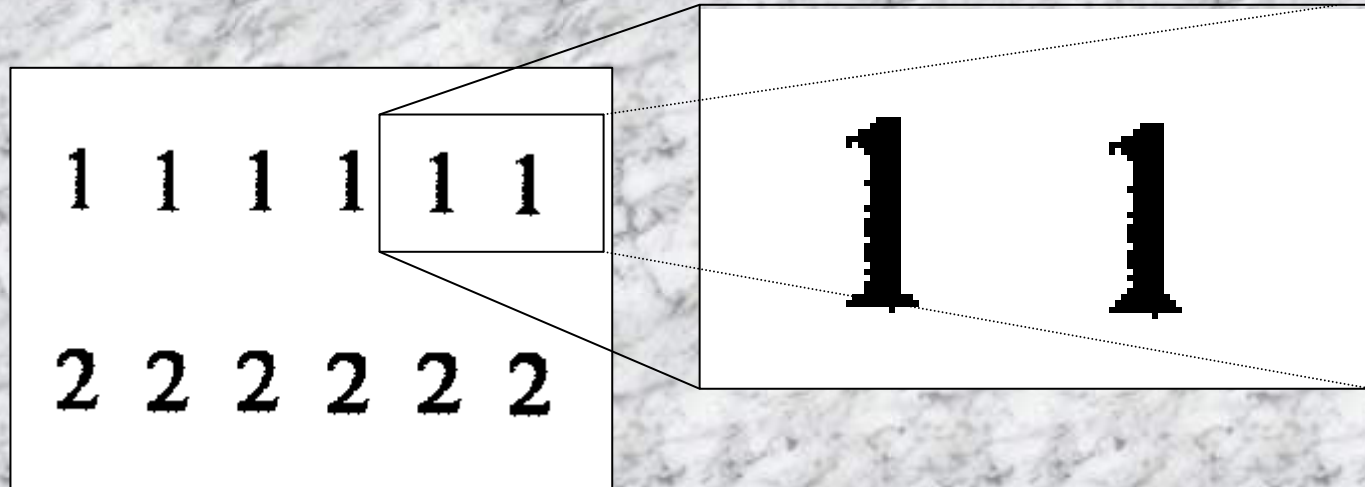
With Anchor



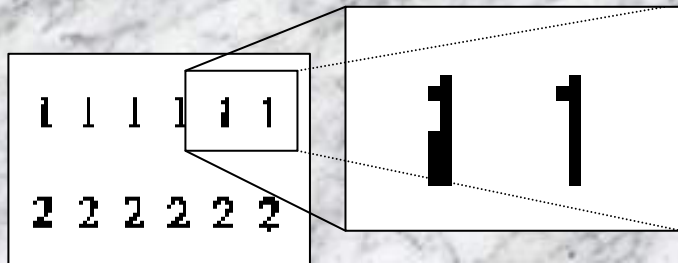
Shrink



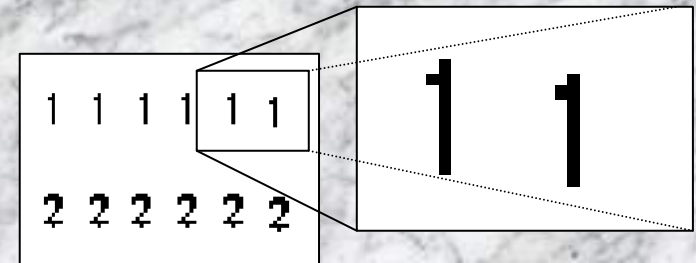
Anchoring - Example



**Original Image with
Identical Characters**

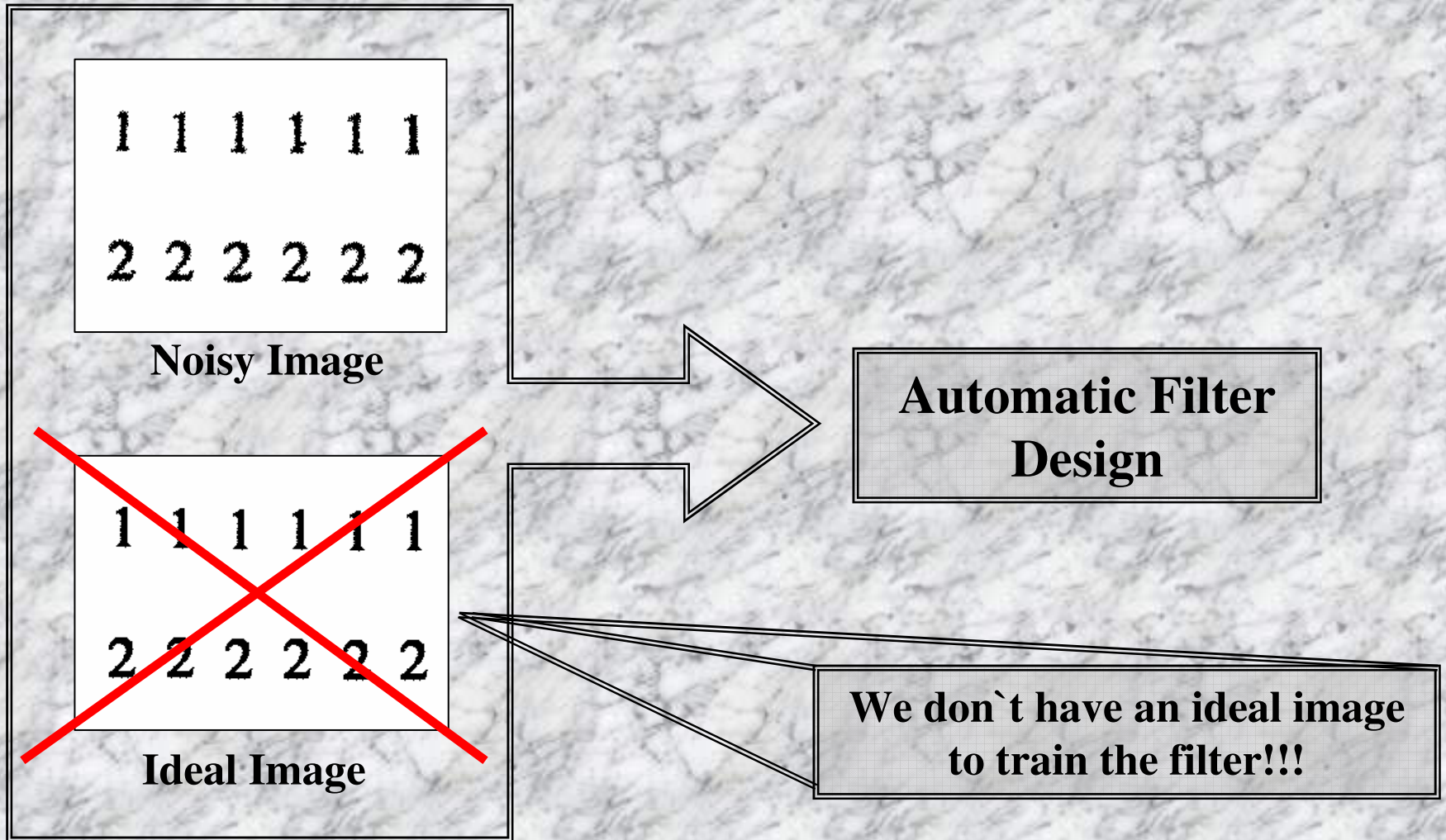


**Direct Sampling
(doubled scale)**

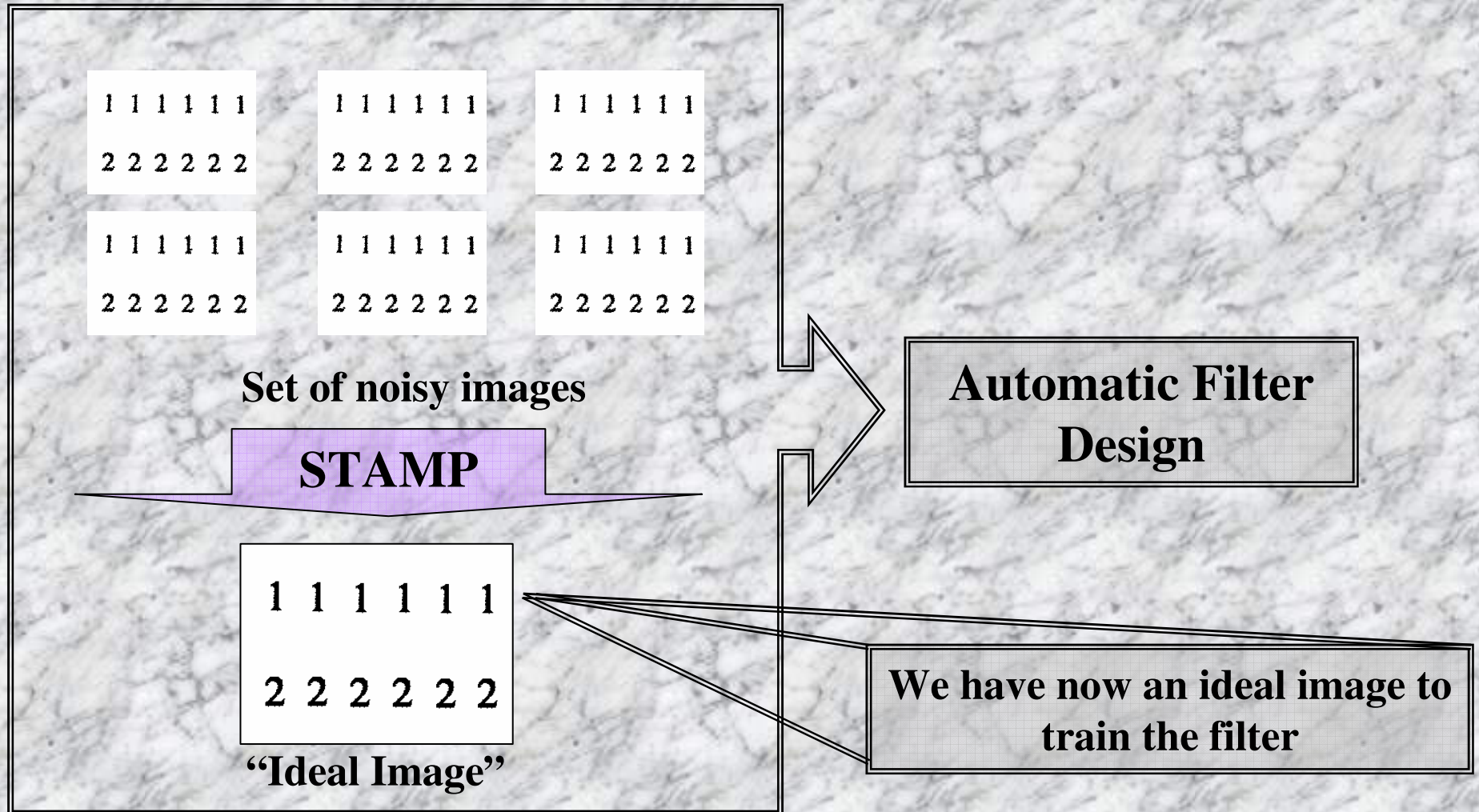


**Sampling after
anchoring (doubled
scale)**

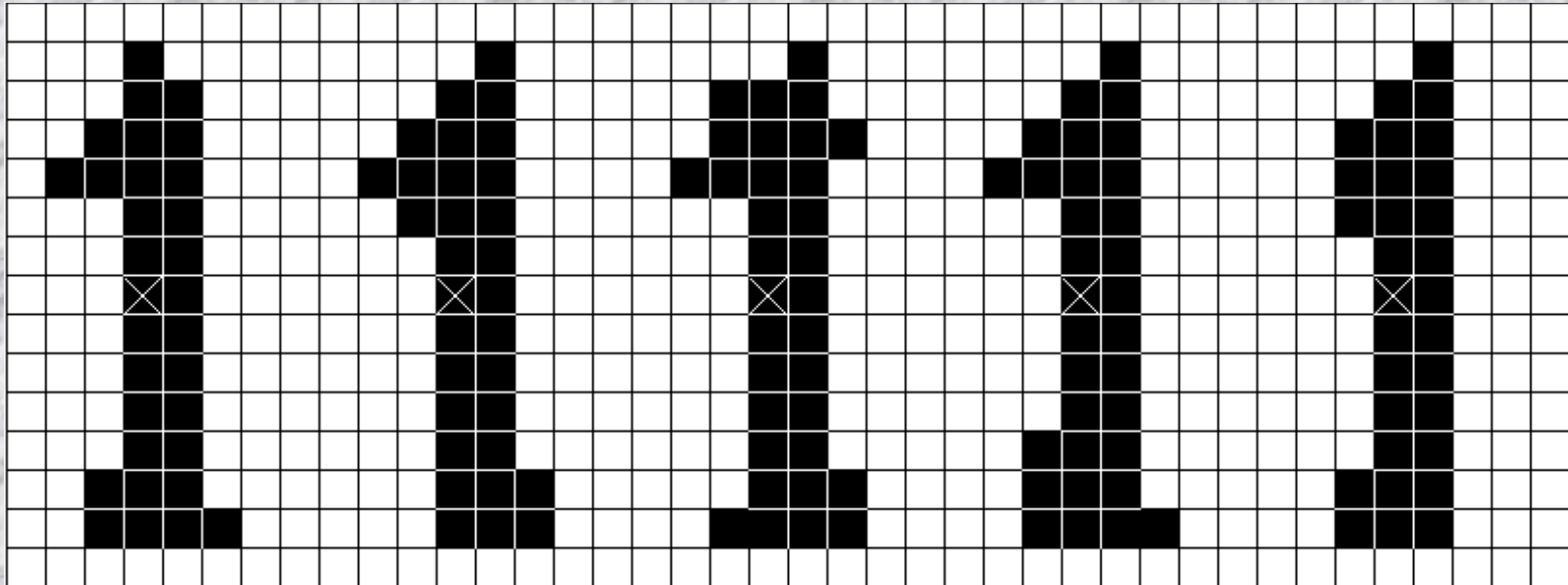
Edge noise filtering



Edge noise filtering - Stamp

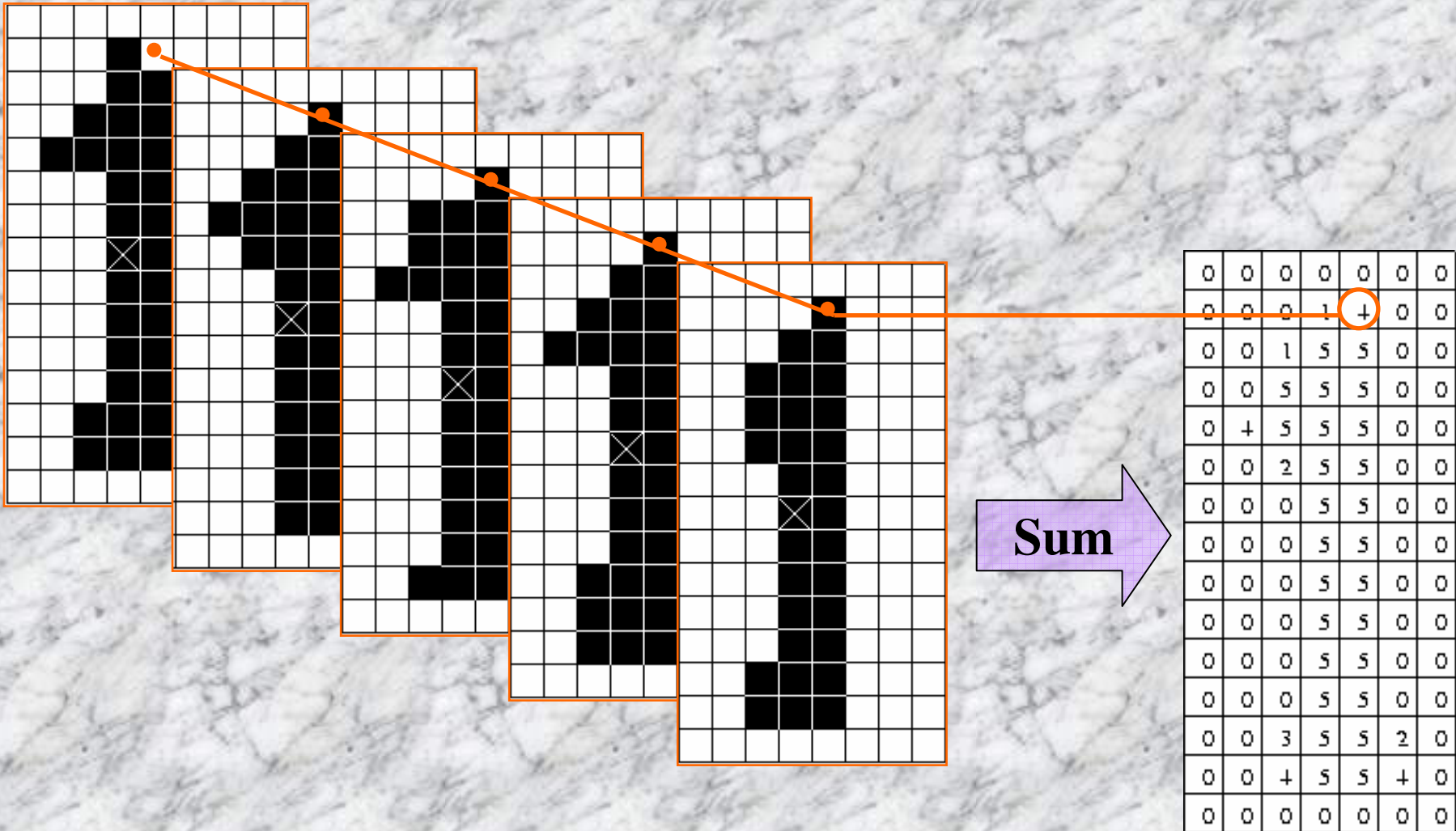


Edge noise filtering - Stamp

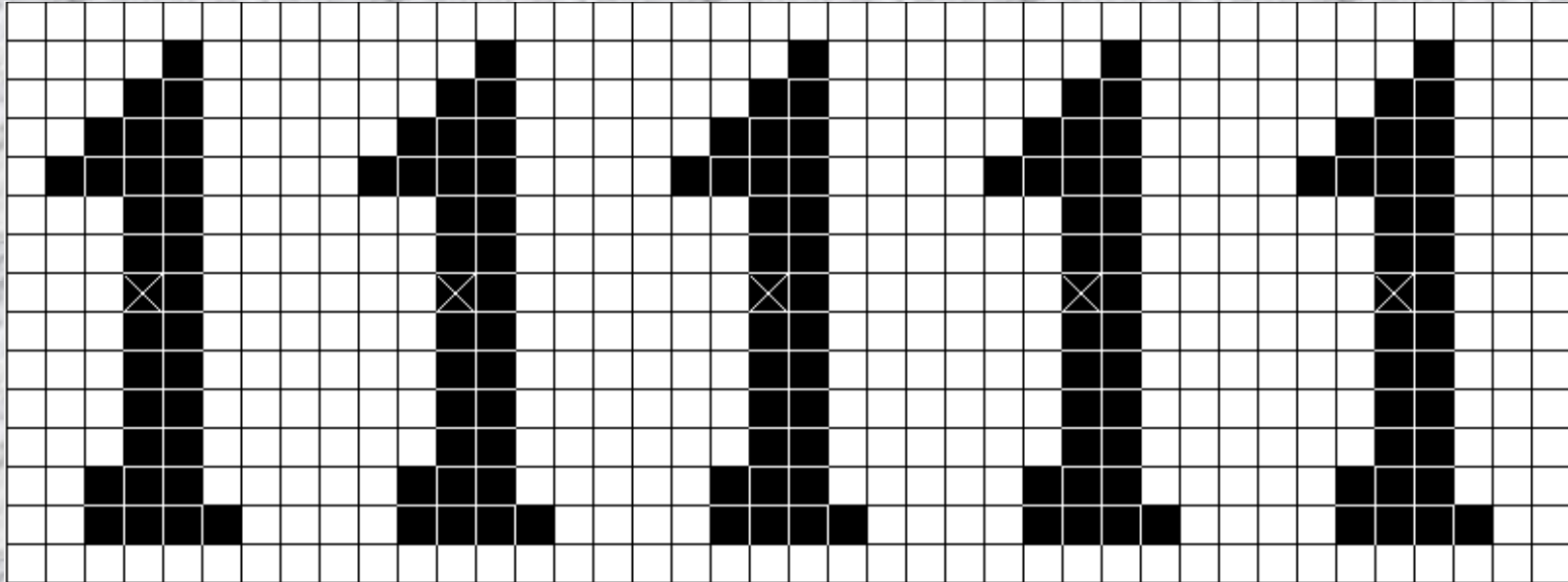


- 5 edge noise realizations of the number "one"
- From them we want to obtain a good almost "ideal" number "one"
- For this, we mark the mass center of each digit (here shown with a cross)

Edge noise filtering - Stamp

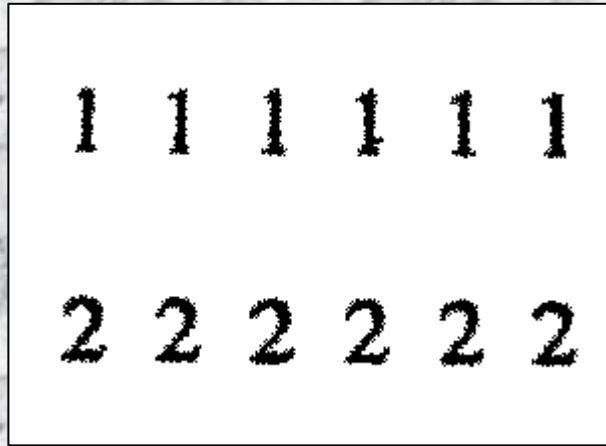


Edge noise filtering - Stamp

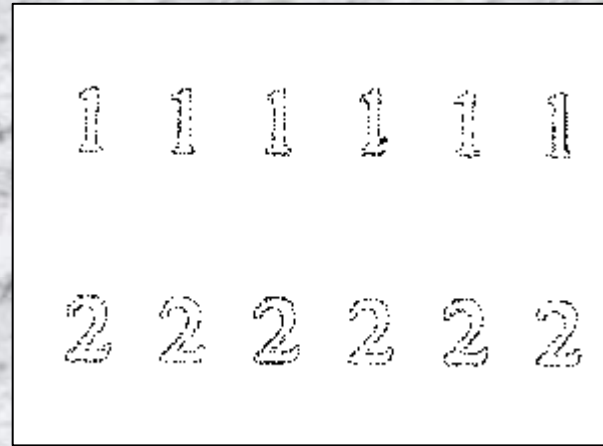


- Each digit is replaced by the "ideal" digit, matching the centers (cross)
- With this image and the noisy original image, we have the training pair to automatically project the filter

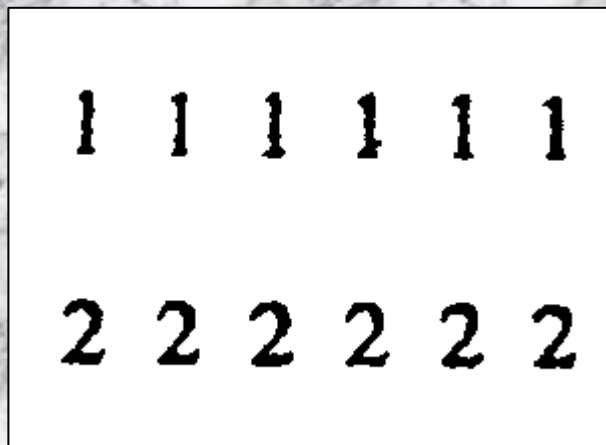
Edge noise filtering - Example



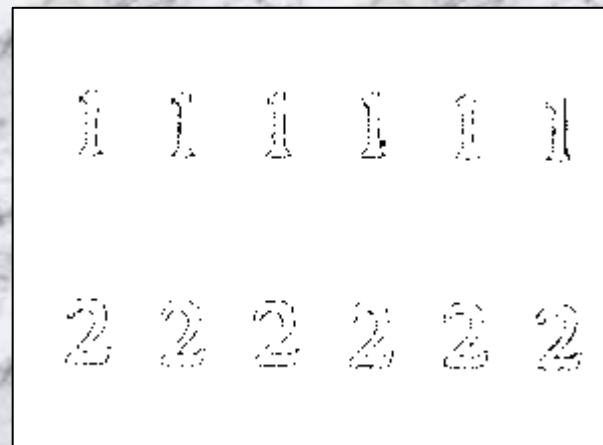
Original Image



Differences with ideal stamped image

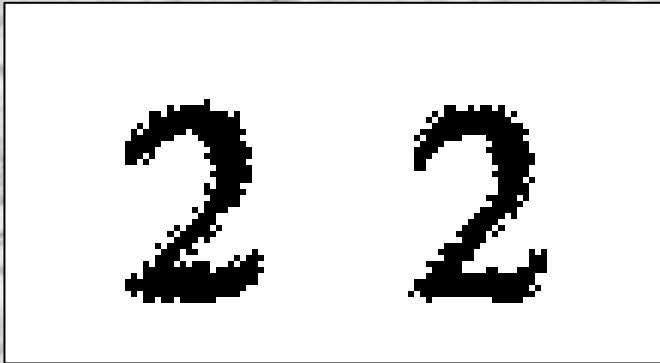


Filtered Image

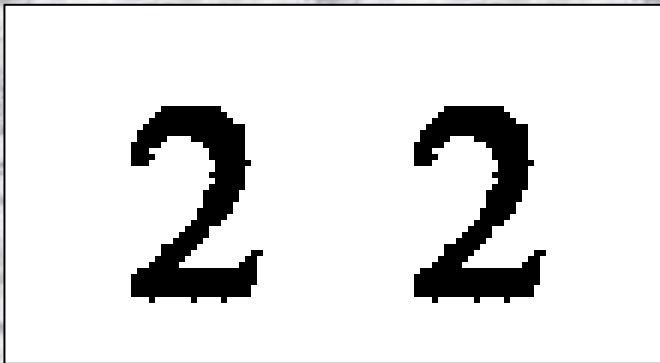


Differences with ideal stamped image

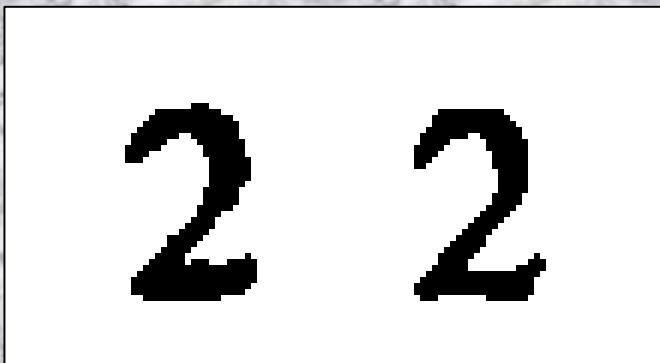
Edge noise filtering - Example



Detail of original
Image

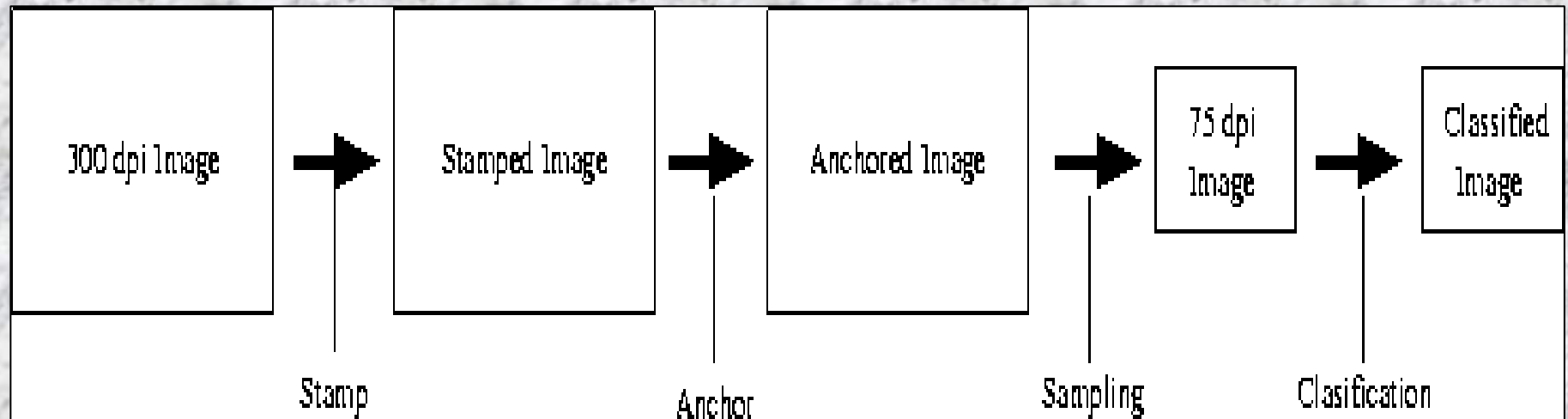


Detail of ideal stamp



Detail of application of
the stamp based noise
edge filter

Full sequence

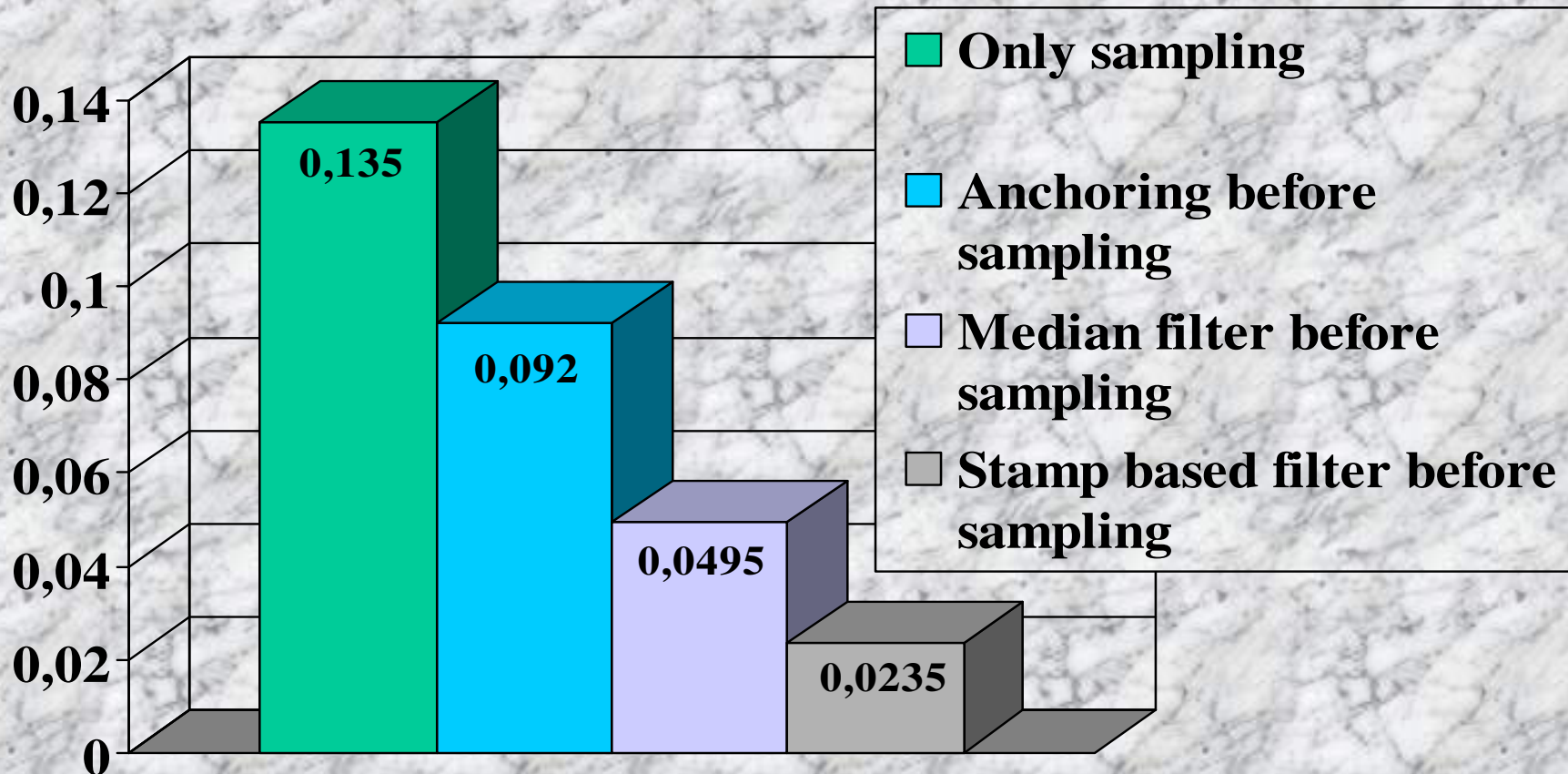


Normalisation Diagram

- ◆ The complexity grows with the windows W
- ◆ Anchoring \Rightarrow Decrease image resolution
- ◆ Stamp \Rightarrow Normalise characters shape

Experimental Results

Errors Graphs (%)



Conclusion

- **We changed the problem of designing sophisticated shape classifier into one of designing simple classifiers for filtered images**
- **Correct classification were from 86.5% to 90.8% using anchoring thecnique**
- **After filtering, the values dropped to about 97%**
- **The goal of the filter is character normalization, and this is exactly modeled by indempotent operators**
- **The next priority of our research on this subject is the development of a technique for the design of indempotent operator**

Aknowlodgements

- **The authors have received partial support from**
 - **FAPESP and**
 - **CNPq.**
 - **The authors also thanks the students**
 - **Teofilo E. Campos**
 - **Rogério Feris**
 - **Archias A. de A. Filho**
 - **Franklin C. Flores and**
 - **Fabiano C. Sousa**
- that have developed part of the software used in the paper**