Detección automática de neuronas usando aproximaciones de Machine Learning

Gadea Mata Martínez

Seminario de Informática “Mirian Andrés”
16 de mayo de 2016

Rotterdam (Países Bajos)

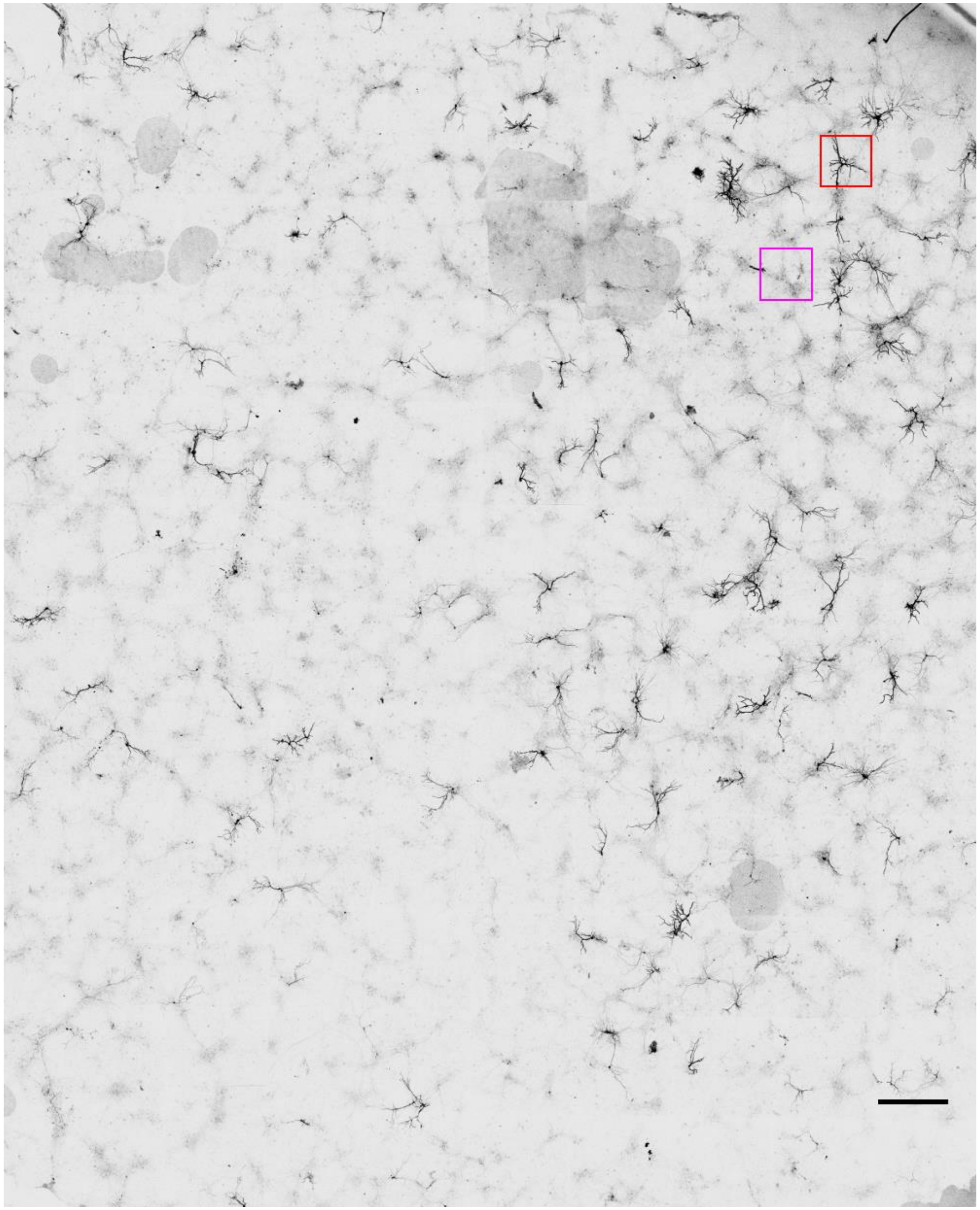
Find neurons

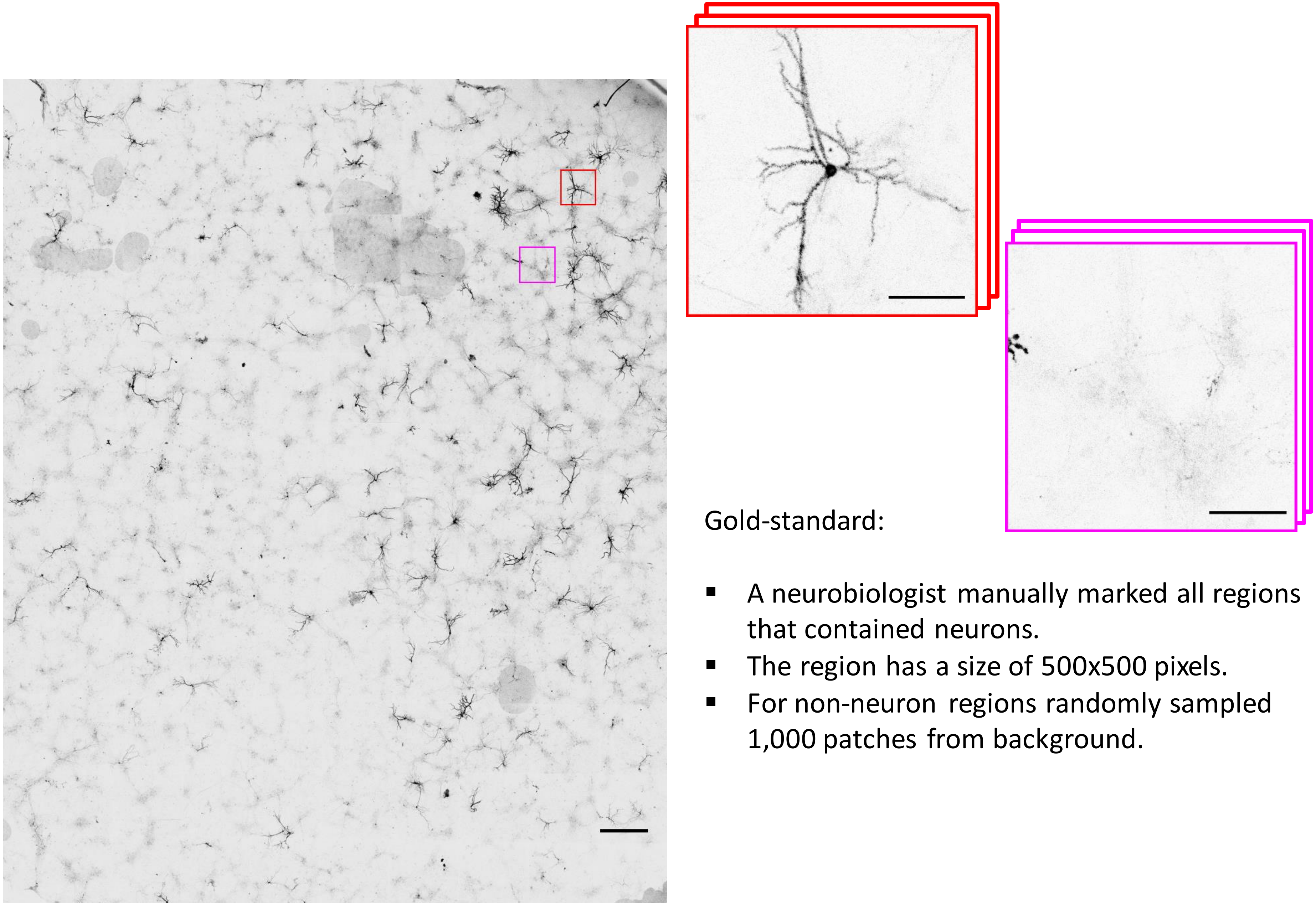
Recognize them

+

Localize them

- Image acquired with a Leica SP5 Confocal Microscope (20x lens).
- The size is about 10,000x12,000 pixels (about 70mm^2).
- One mosaic contain on the order of 40 transfected (with protein GFP) neurons.
- We have studied 8 images (mosaics).





Gold-standard:

- A neurobiologist manually marked all regions that contained neurons.
- The region has a size of 500x500 pixels.
- For non-neuron regions randomly sampled 1,000 patches from background.

We considered **two machine-learning approaches.**

1. WNDCHRM which computes an extensive list of image features.
   Over 1,000 features were computed among them:
   - different types of polynomial decompositions
   - high contrast features
   - pixel statistics
   - texture descriptors (*)

Not all features are equally relevant. The programme computes a Fisher score for each, which allows building a ranked preferences of features. And it only works with the top-15% features according this score.

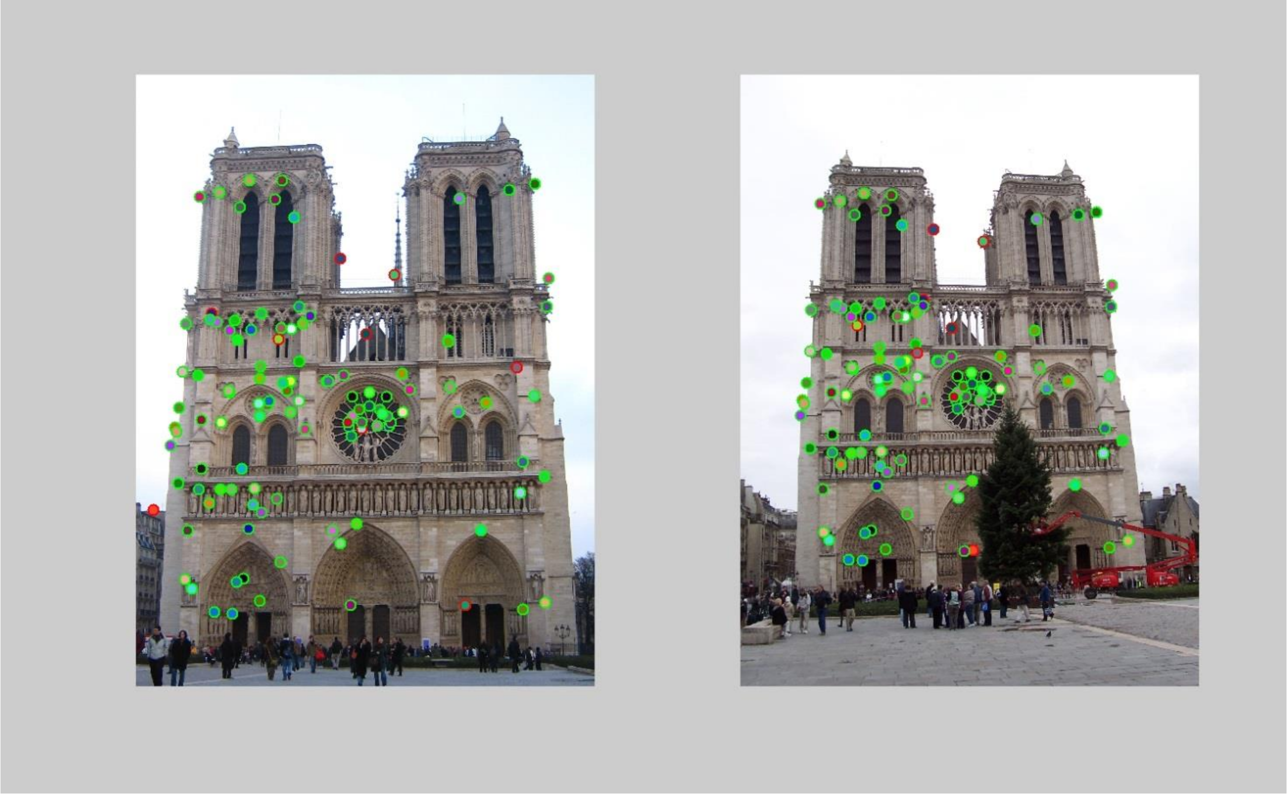
Classification of patches was done using a **weighted neighbor distances** (WND) **classifier** (it's a version of kNN algorithm).

(*) An **image texture** is a set of metrics calculated in image processing designed to quantify the perceived texture of an image. Image texture gives us information about *the spatial arrangement of color or intensities* in an image or selected region of an image.
Image textures can be artificially created or found in natural scenes captured in an image. Image textures are one way that can be used to help in segmentation or classification of images.

Study of images

We considered two machine-learning approaches.

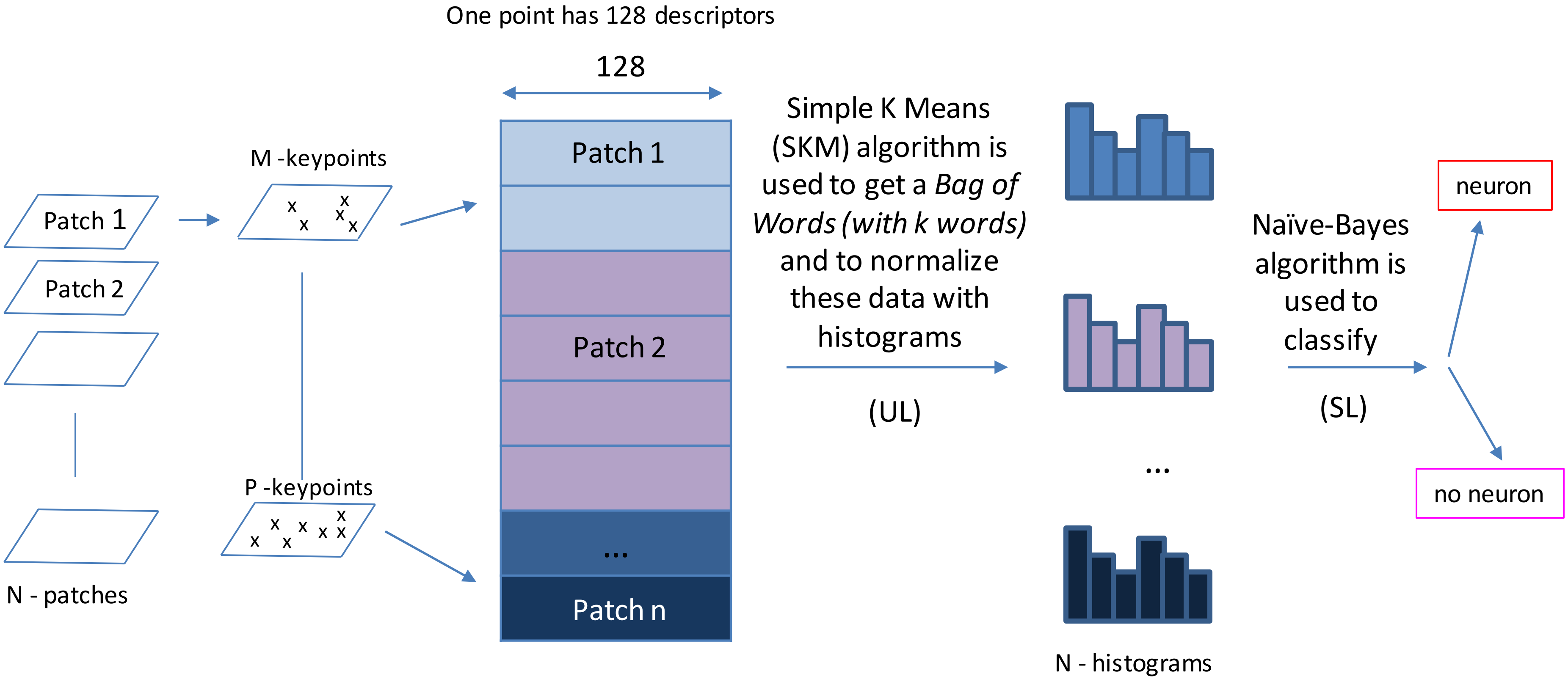
2. Computing the scale-invariant feature transform (SIFT)
   This algorithm find the keypoint descriptors of the image that are:
   - highly distinctive,
   - invariant to image scaling, rotation, a range of affine distortions,
   - robust to noise and illumination change.



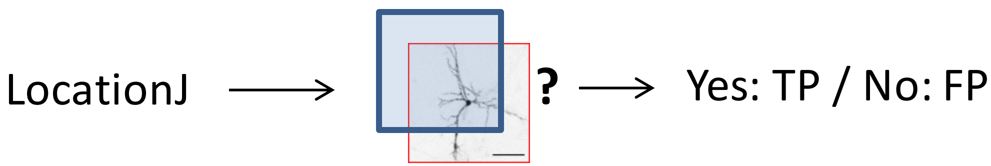
Study of images

2. Computing the scale-invariant feature transform (SIFT)

To classify the patches based on these features we used:
- a bag-of-words model
- Naïve-Bayes (NB) classifier from the WEKA toolkit.

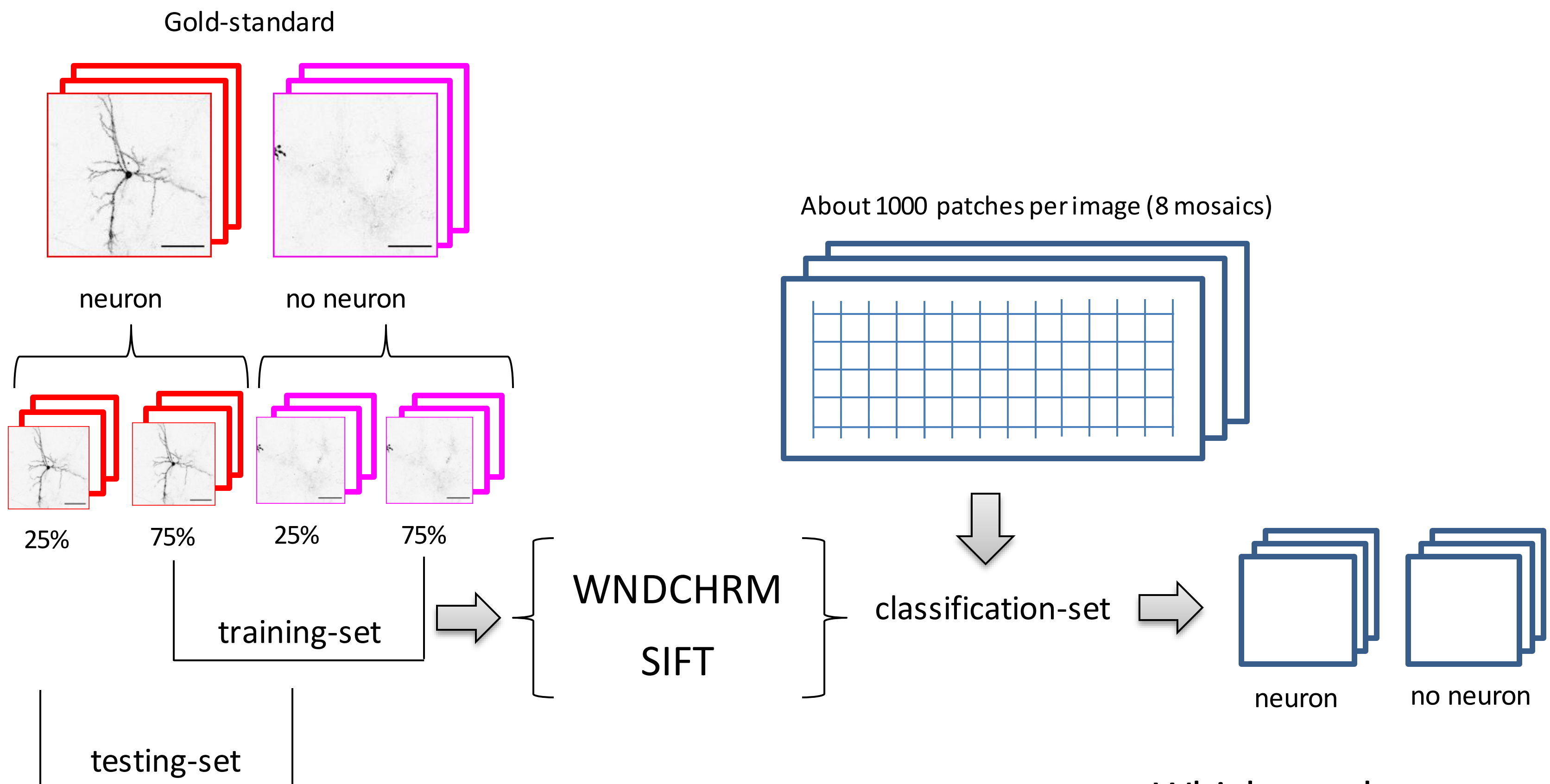


Analysis of images:



method	recall	precision
LocationJ	0.80	0.86
WNDCHRM-A	1	0.62
SIFT-A	0.74	0.46

Supervised:



Which are the true-positives?

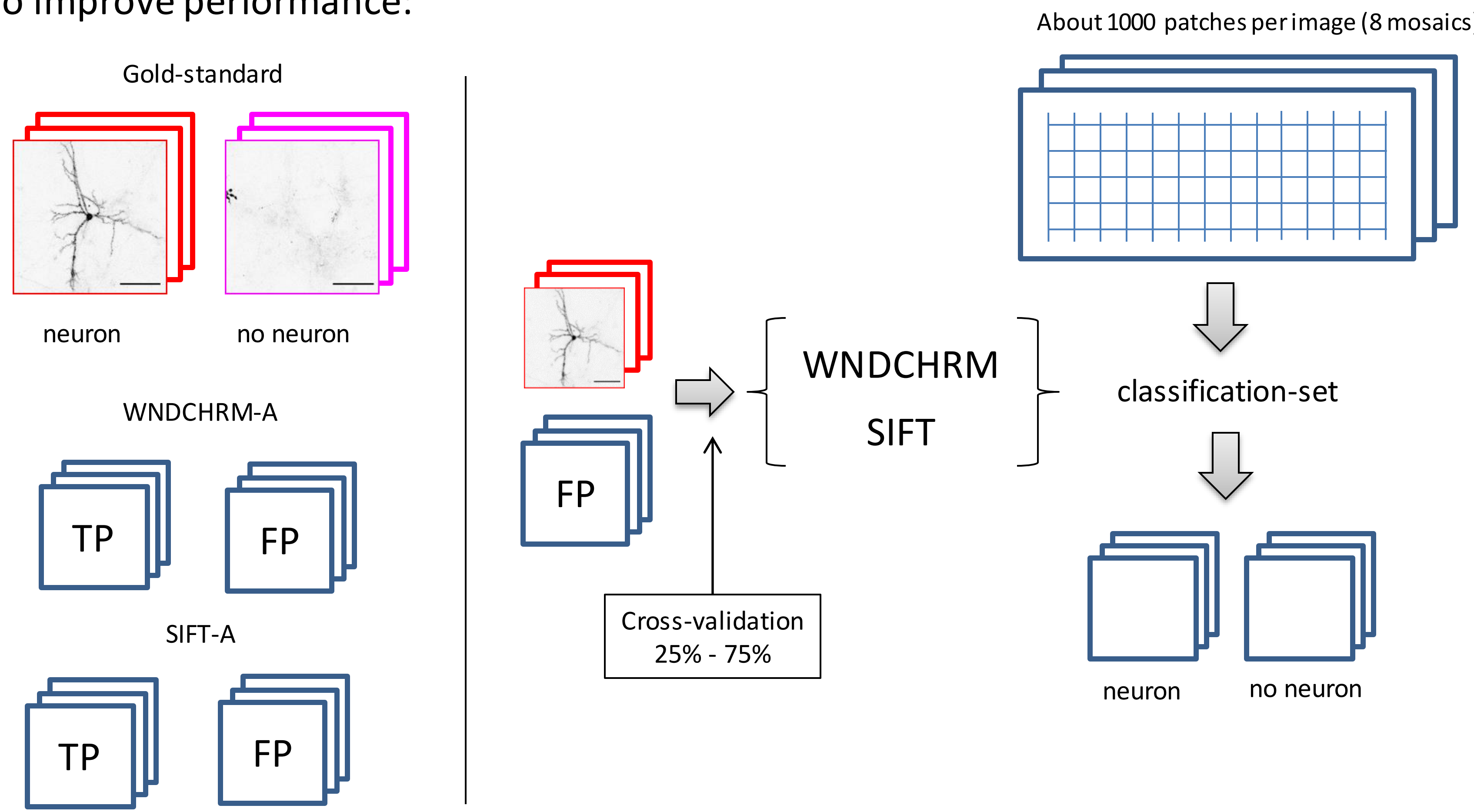
P: positive | TP: true positive | FP: false positive | FN: false negative | recall = TP/P | precision = TP/(TP+FP)

Results:

method	recall	precision
LocationJ	0.64	0.50
WNDCHRM-A	0.89	0.19
SIFT-A	0.60	0.26
WNDCHRM-B	0.66	0.44
SIFT-B	0.94	0.57

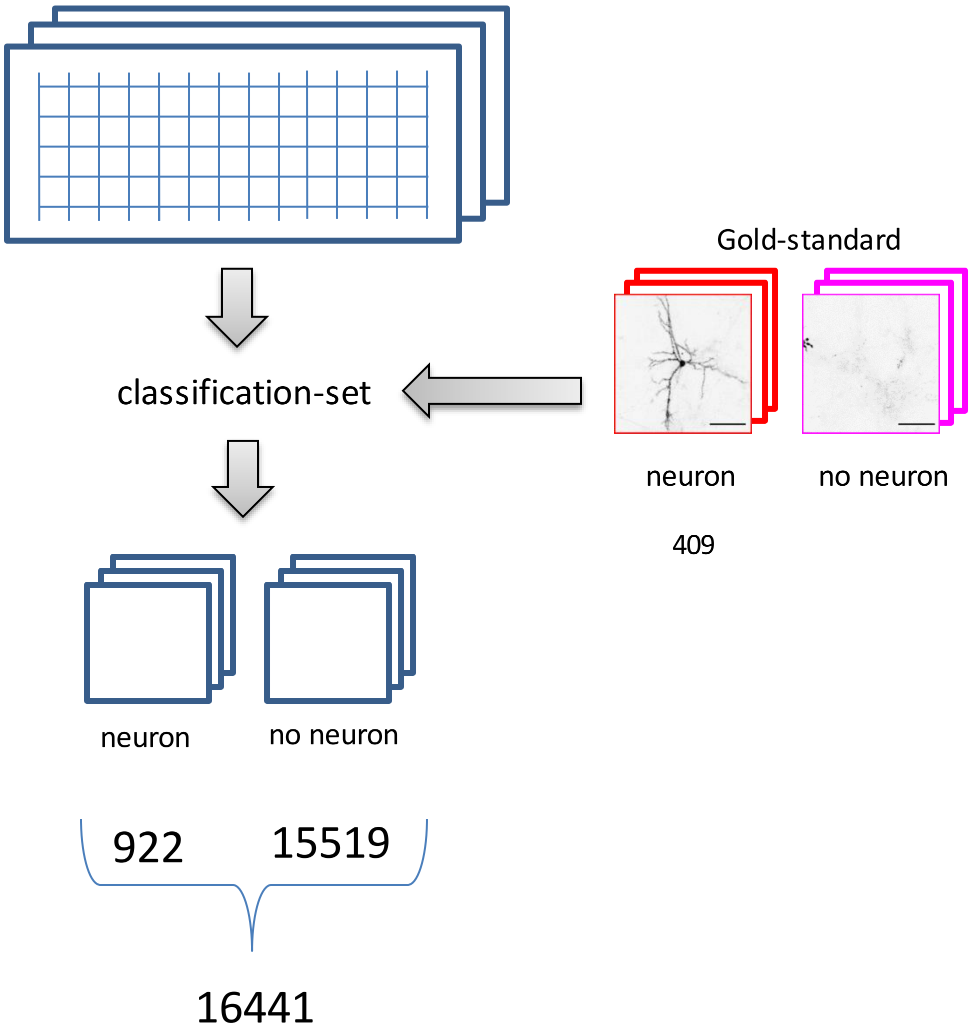
International Symposium on Biomedical Imaging (ISBI) Praga 2016

To improve performance:



P: positive | TP: true positive | FP: false positive | FN: false negative | recall = TP/P | precision = TP/(TP+FP)

Rotterdam 2016:



SETTINGS:

500x500 pixels

Overlapping 50%

To consider neuron, an overlapping 50%

Settings to calculate the patches

Mosaic nº 6:

Time:	<1’	1’	3’	5’	5’	6’	29’
Over. % grid	25	50	70	75	75	80	90
Over. % pos.	25	50	70	25	75	80	90
Nº Pos patches	298	253	233	2688	236	225	221
Nº patches	750	1665	4650	6660	6660	10304	41216

- We work with 8 mosaics with different sizes. The time is similar for each one, except to mosaic number 4 that the time to obtain the patches is 16 minutes.

- Colaboración con Carlos Fernández (Universidad de A Coruña)

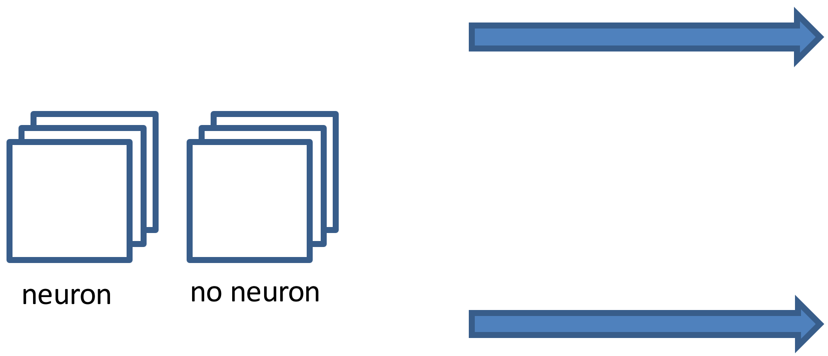
Obtener las características de los patches de los mosaicos

↓

Estudiar diferentes algoritmos de Machine Learning

↓

Analizar los datos para ver que algoritmo es mejor



WNDCHRM
Get the features (1059)

SIFT
Normalize the features, get histograms

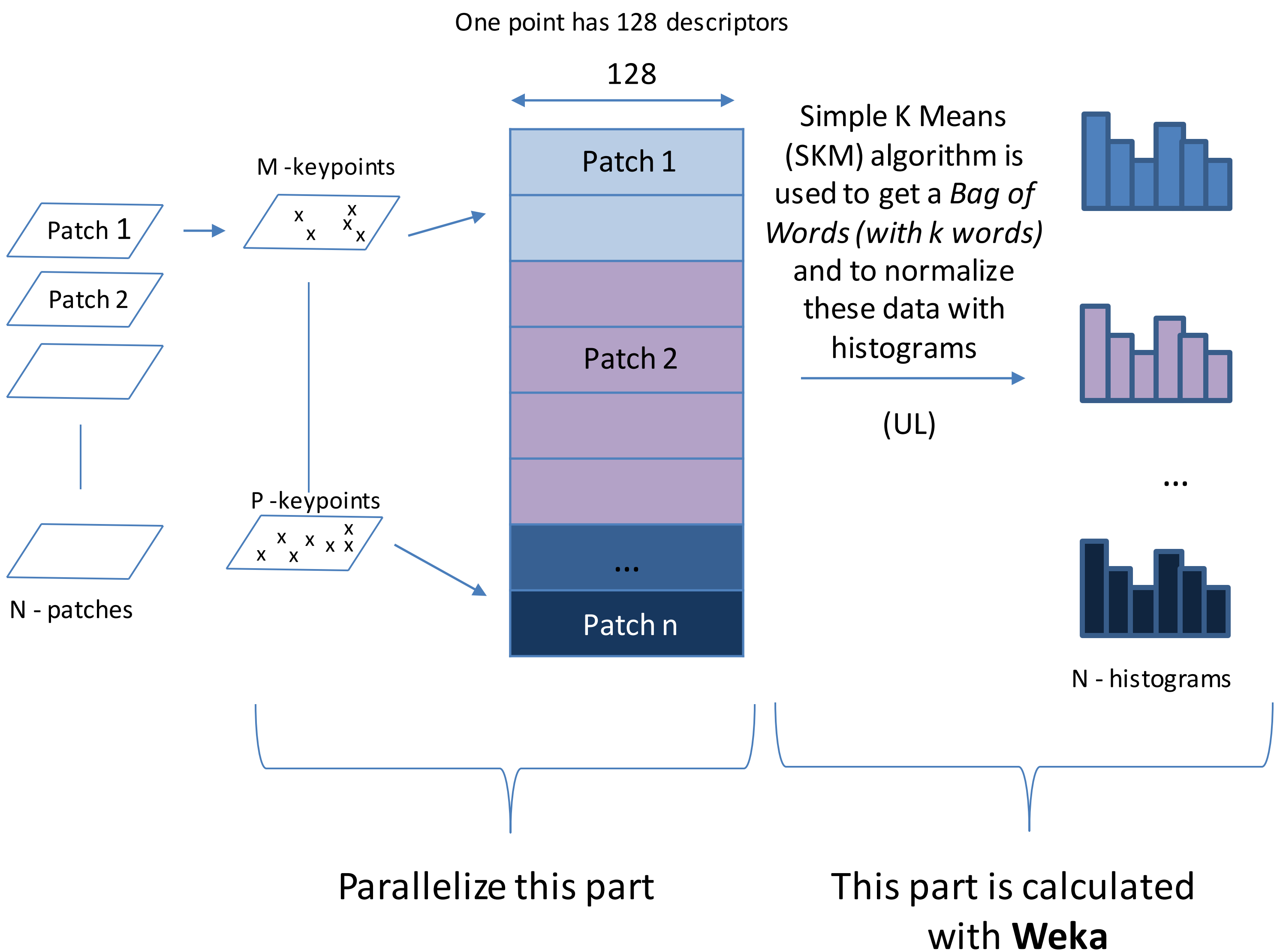
Technical Problems:

- Time
  - WNDCHRM: all mosaics 48h

m01	6’	m05	6’
m02	6’	m06	8’
m03	3’	m07	8’
m04	44’	m08	6’

  - Obtain histograms: Study **a mosaic** with 100 words takes +24h

SIFT features:



Now, it takes 17h (a mosaic (m06) with 100 words)

Mosaic m05	Nº of words	Time
Get histograms	20	48’
	40	43’
	60	56’
	80	65’
Total process	80	3h

The main problem is the time.

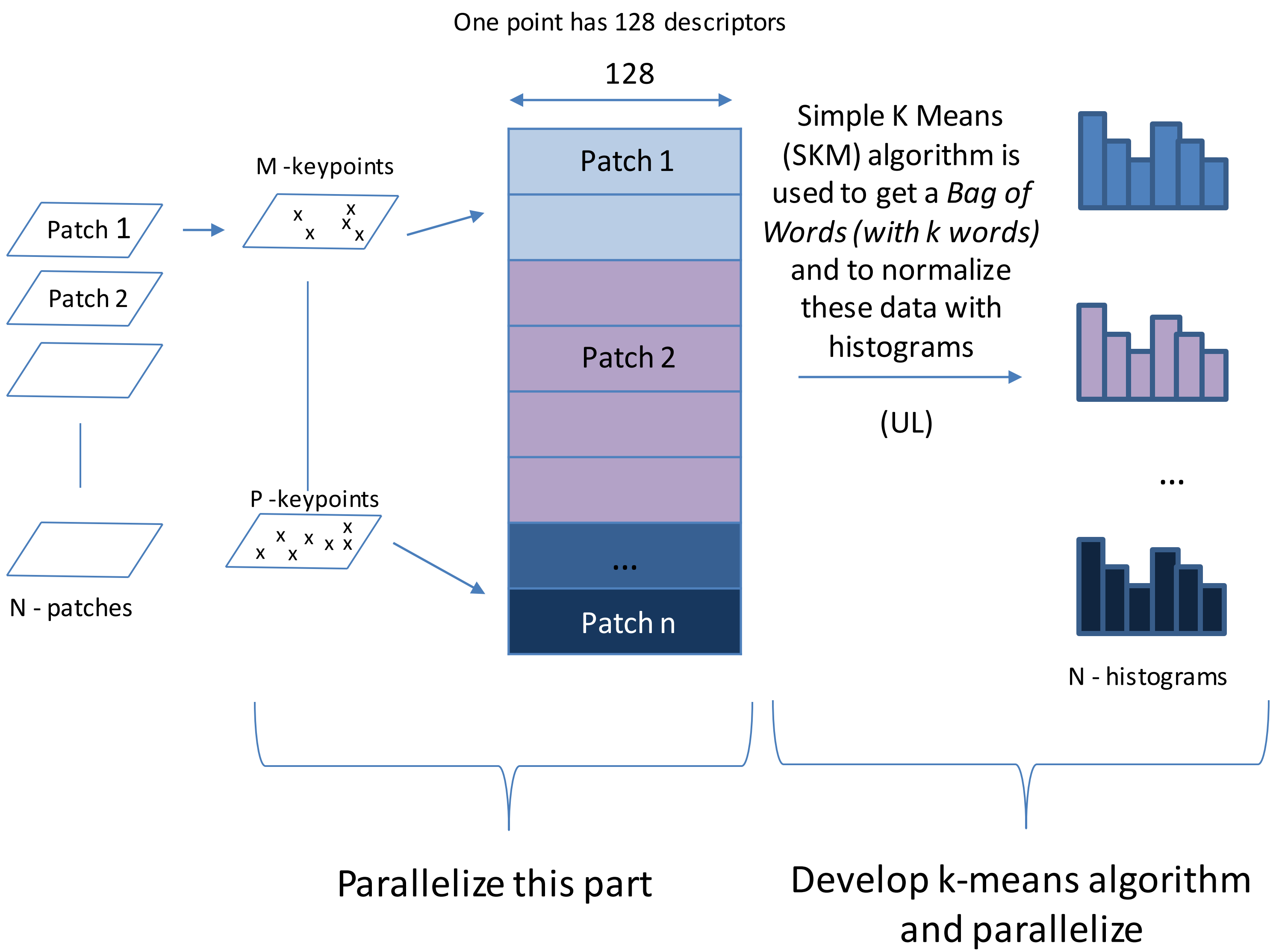
This is only with one mosaic. However we need to work with all mosaics to the same time.

The number of sift features in all mosaics is **743.907**

Settings:

- Size of patch
- Overlap % grid
- Overlap % to consider positive a patch
- Nº of words

SIFT features:



Under construction

Conclusion:

- Have a system to calculate this features (WNDCHRM and SIFT)
- Obtain the data and study them (Carlos Fernández)
- Analyze the results
- Compare them with other methods

Thanks! Dank u wel! Gracias!