Creatis

HEALTHGRID 2010 CONFERENCE PARIS

Mean-Shift scale parameters optimization on the EGEE grid

Ting LI, **Sorina CAMARASU-POP**, **Tristan GLATARD**, **Thomas GRENIER**, **Hugues BENOIT-CATTIN**

CREATIS, Lyon, FRANCE

29/06/2010

UNIVERSITÉ DE LYON

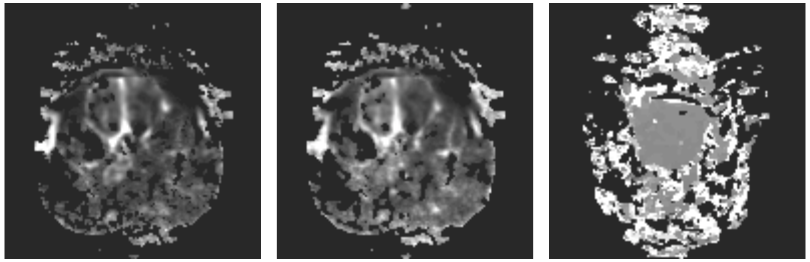
cnrs dépasser les frontières

Inserm Institut national de la santé et de la recherche médicale

Medical Context

--- clustering is a powerful tool to analyze multi-dimensional data

- Images and parameters
  - Multi-modality: Ultrasound, PET, MRI, X-ray, etc.



cbf **peak** **ttp**

MRI parameters for penumbra segmentation in ischemia [M.Wiart]

clustering →

- Characterize pathology
- Segmentation
- Detect disease

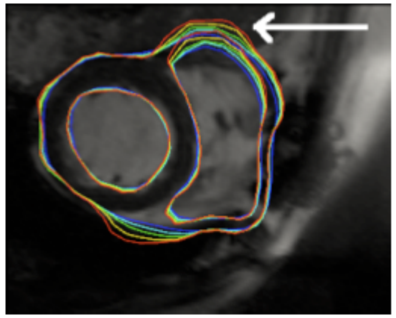
- Clustering
  - Filtering: noise removing, edges preserving, etc.
  - Segmentation: cluster or separation of clusters

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

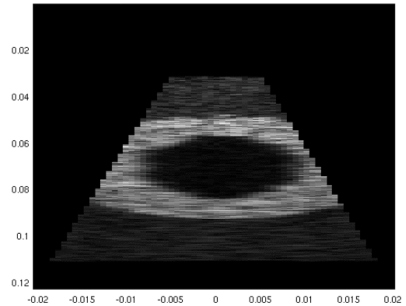
Grid computing for medical imaging

--- used in production for a wide range of applications

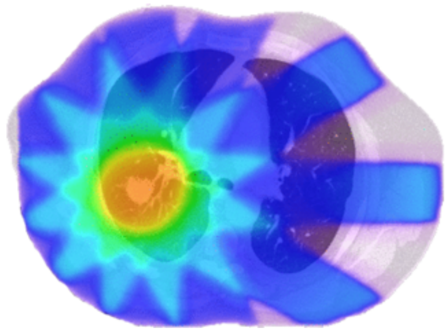
- Parameter sweeps
  - Functional MRI [Olabarriaga *et al,* Biogrid'07]
  - Image segmentation [Maheshwari *et al*, HG'09]
    [Specovius *et al,* HG'10]
- Data parallelization
  - On image lines / slices / volumes / sequences
  - Pharmacokinetic modelling [Blanquer *et al,* HG'08]
  - Image simulation
- Monte-Carlo simulation
  - Radiotherapy [Maigne *et al,* PPL 2004]
  - Image simulation (PET, CT)



Sweep on segmentation parameters
[Maheshwari ***et al***, HG'09]



Ultrasonic 2D+t simulation



GATE simulation

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

OUTLINE

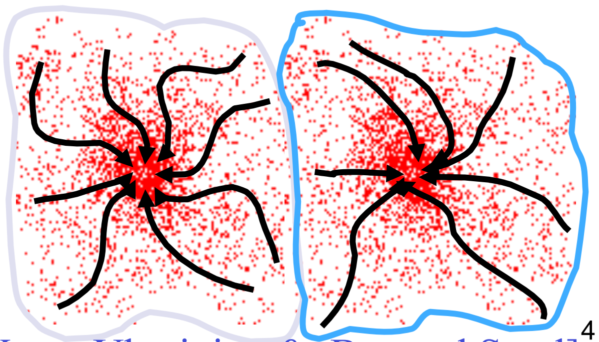
1. Context

2. Mean-Shift presentation

3. Experiments and results on bandwidth optimization

4. Grid deployment and execution

5. Conclusion



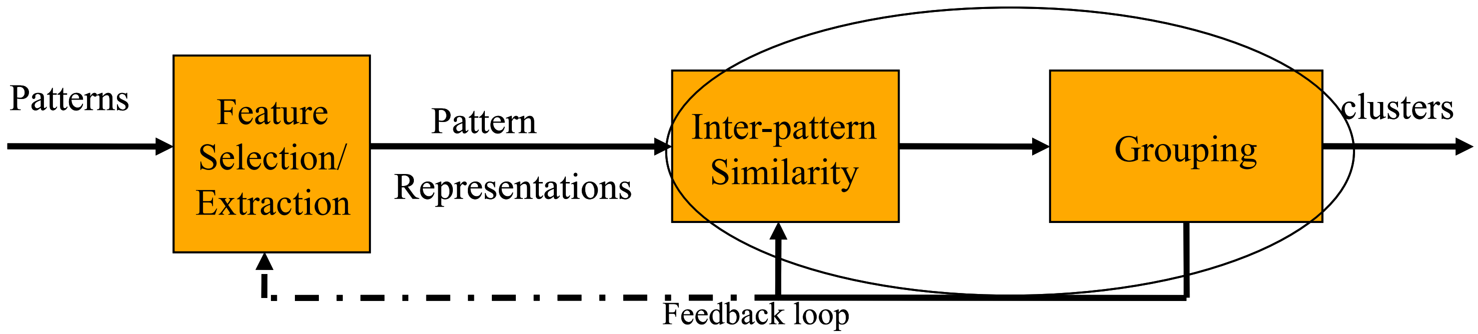
[Yaron Ukrainitz & Bernard Sarel]

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

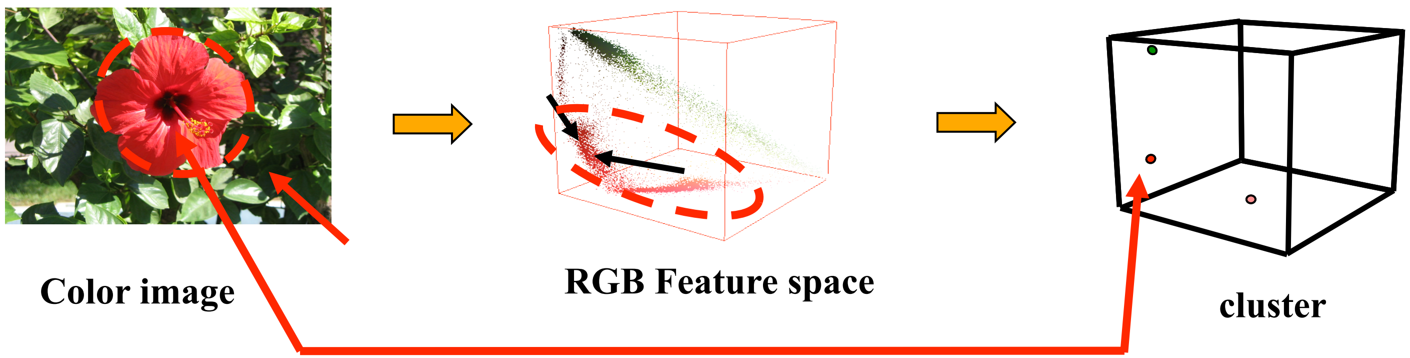
General Clustering Framework

---It is also the mean-shift framework

- General Framework



- Representation in Feature Space

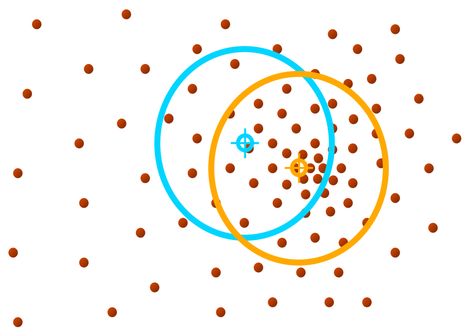


tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Mean-Shift presentation

---Why we use mean-shift as our clustering approach?

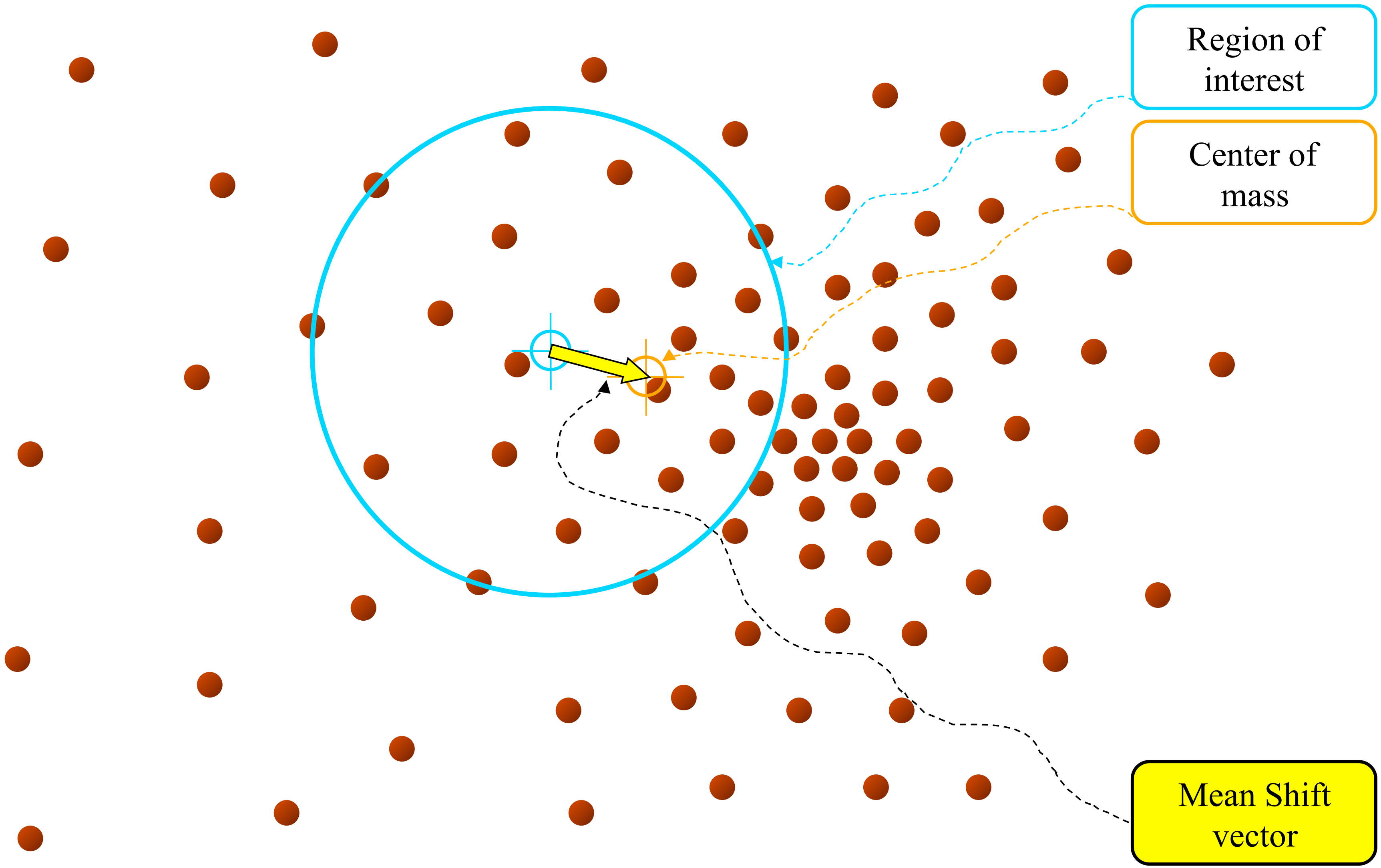
- General Principle
  - Goal: finding modes in a set of data samples
  - Method: probability density gradient of non-parametric estimation
- Characteristics
  - Can handle arbitrary feature spaces
  - Does not assume any prior shape (e.g. elliptical) on data clusters
  - Iterative hill climbing algorithm
- Challenges
  - Finding local optimum (mode)
  - Setting scale parameters H (window size)
  - Time consuming



[Yaron Ukrainitz & Bernard Sarel]

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Intuitive Description



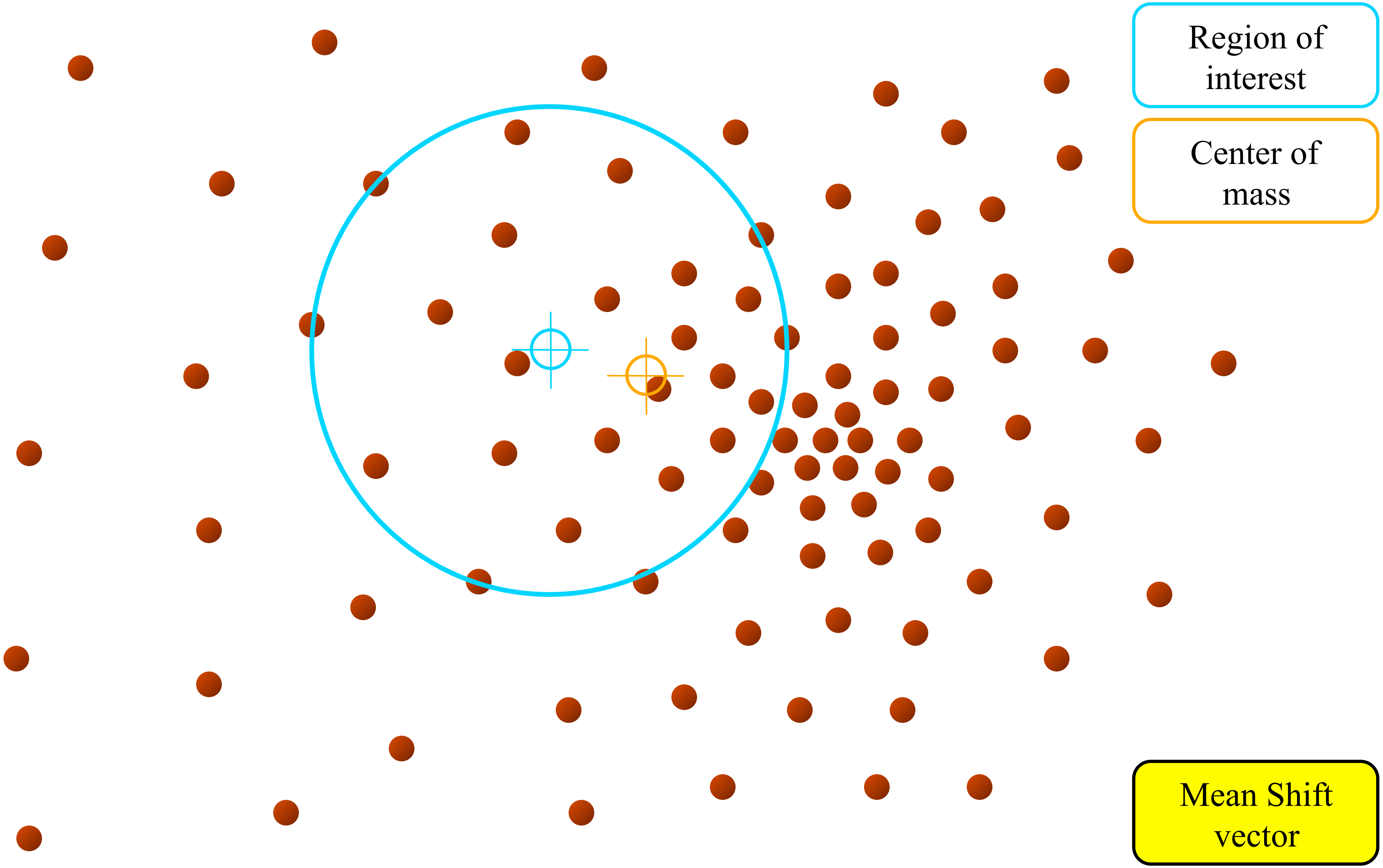
Objective : Find the densest region

Distribution of identical points

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

[Yaron Ukrainitz & Bernard Sarel]

Intuitive Description



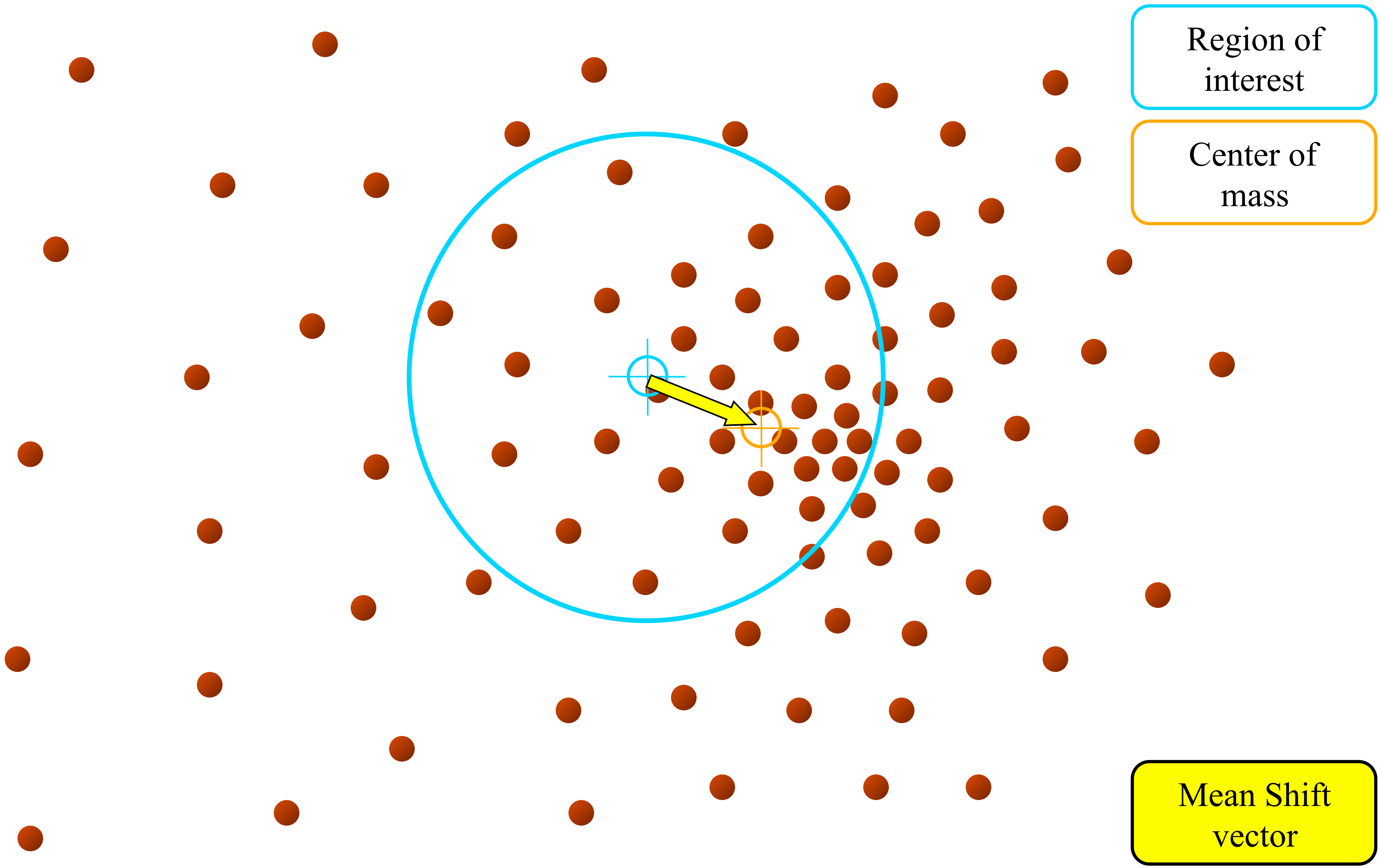
Objective : Find the densest region

Distribution of identical points

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

[Yaron Ukrainitz & Bernard Sarel]

Intuitive Description



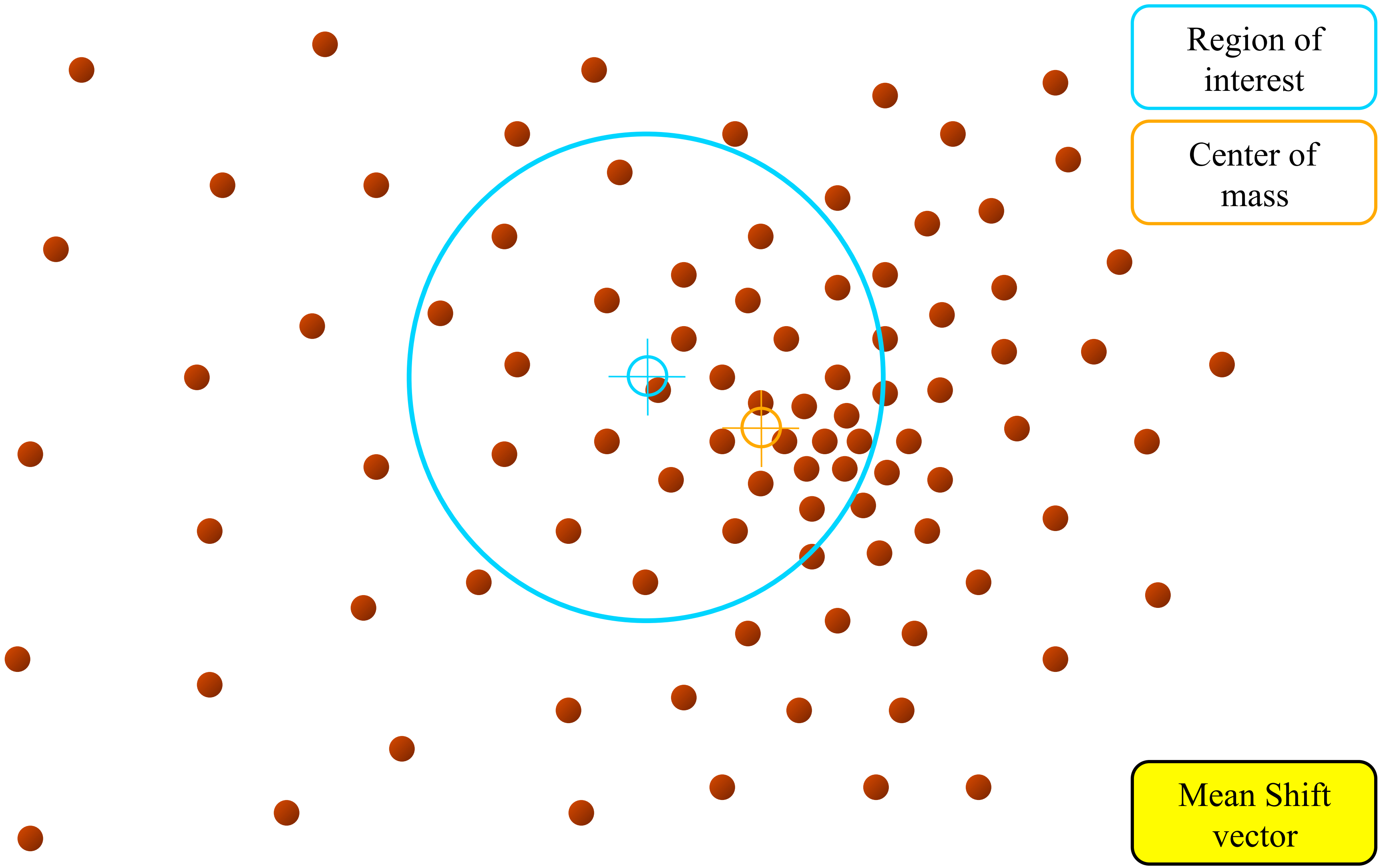
Objective : Find the densest region

Distribution of identical points

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

[Yaron Ukrainitz & Bernard Sarel]

Intuitive Description



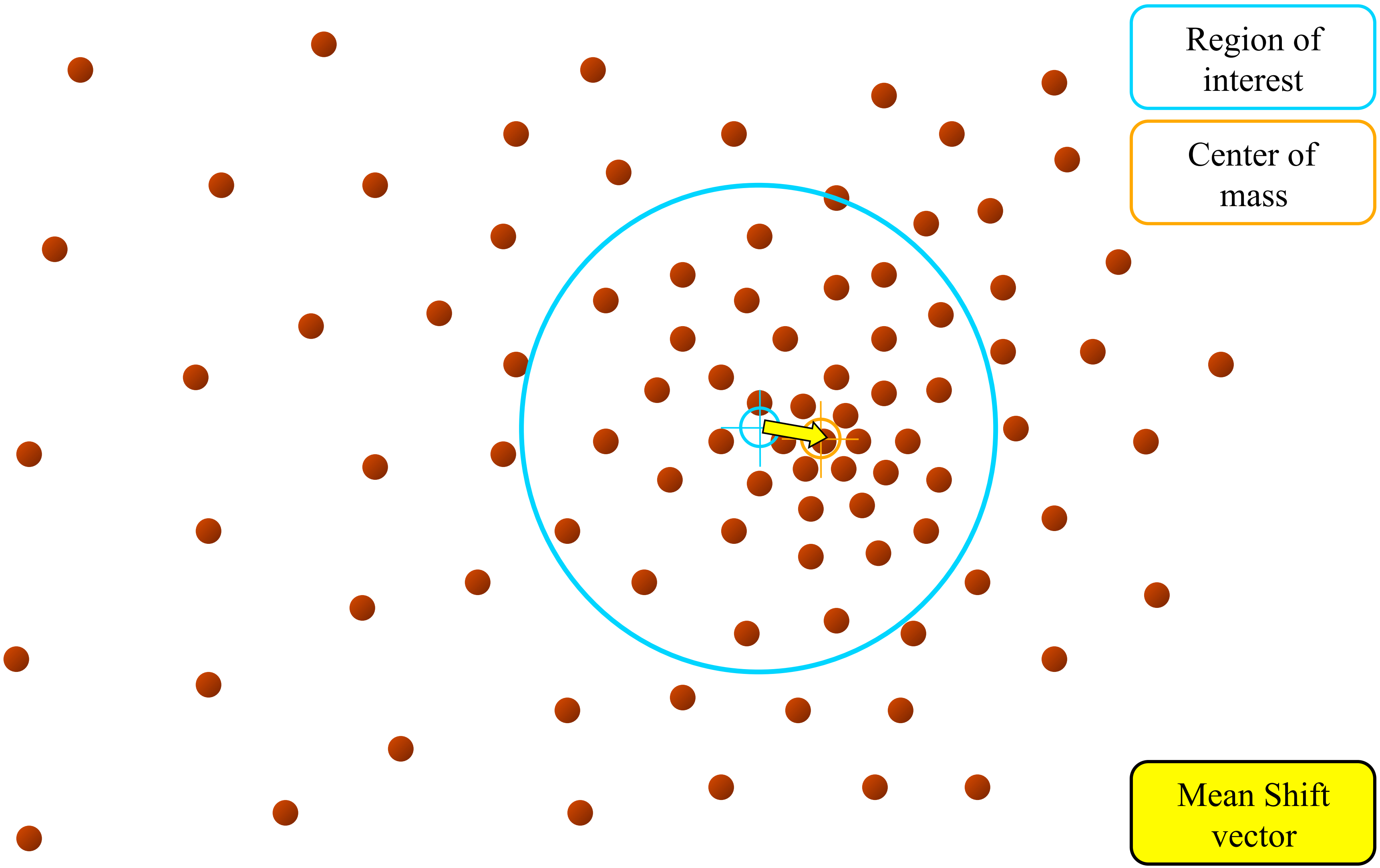
<u>Objective</u> : Find the densest region

Distribution of identical points

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

[Yaron Ukrainitz & Bernard Sarel]

Intuitive Description



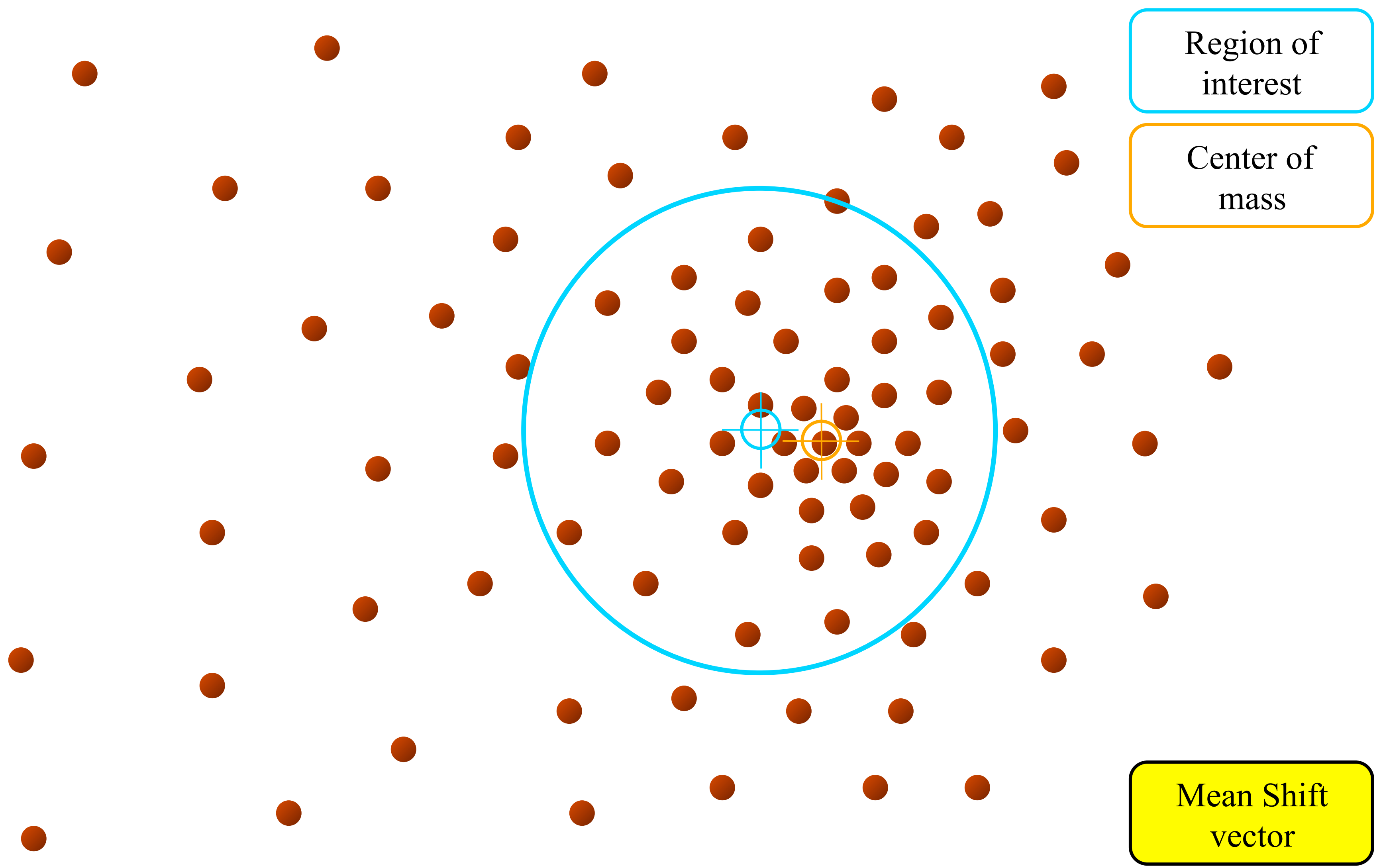
<u>Objective</u> : Find the densest region

Distribution of identical points

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

[Yaron Ukrainitz & Bernard Sarel]

Intuitive Description



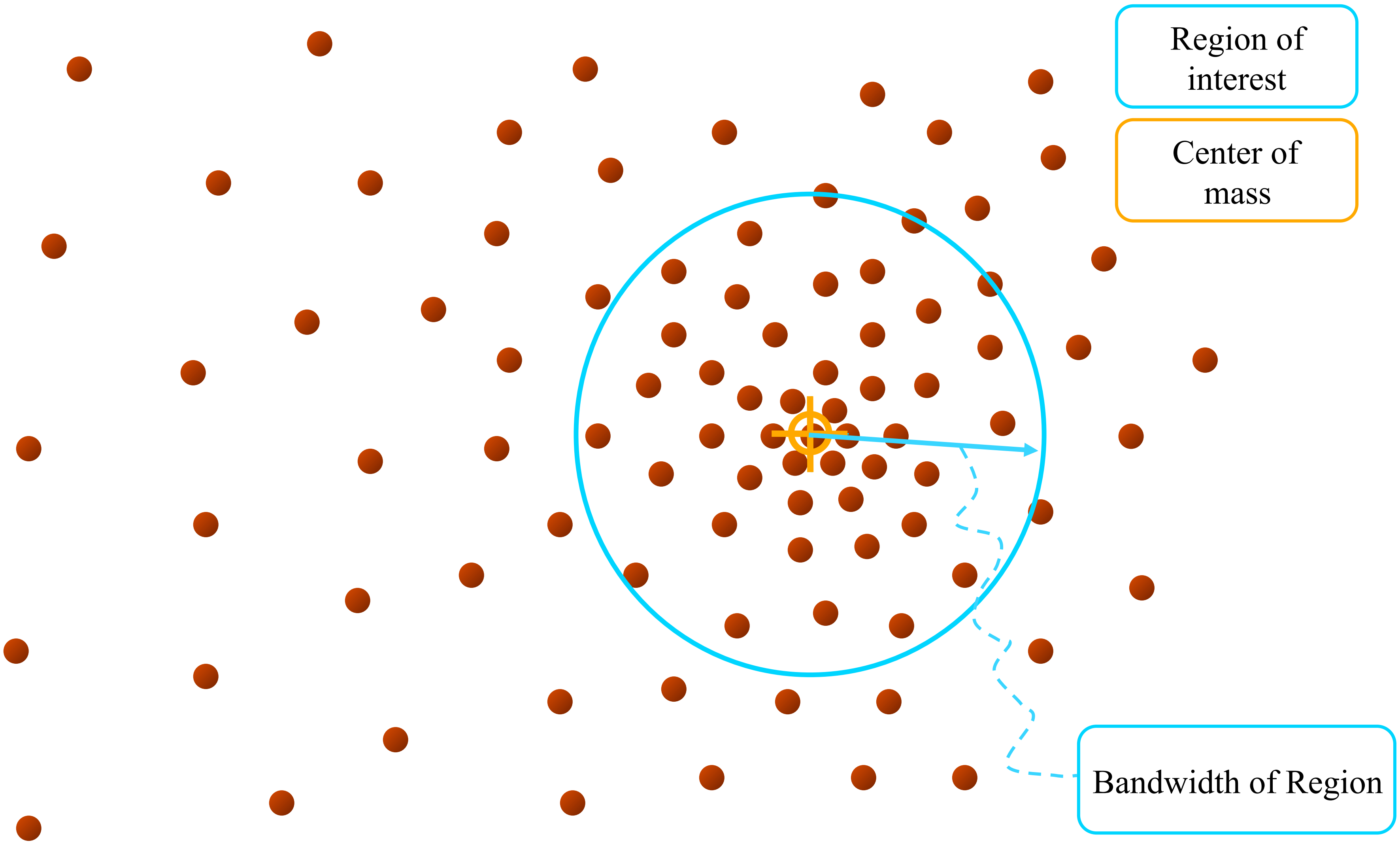
Objective : Find the densest region

Distribution of identical points

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

[Yaron Ukrainitz & Bernard Sarel]

Intuitive Description



Objective : Find the densest region

Distribution of identical points

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

[Yaron Ukrainitz & Bernard Sarel]

Mean-Shift Filtering Procedures

[Comaniciu 99']

- 1. For a given pixel

$$\mathbf{v}^{[t]} = \begin{bmatrix} x \\ y \\ r \\ g \\ b \end{bmatrix}$$

x, y: Spatial 's'

r, g, b: color information

$$\mathbf{v}^{[t]} \in \mathbf{R}^d$$

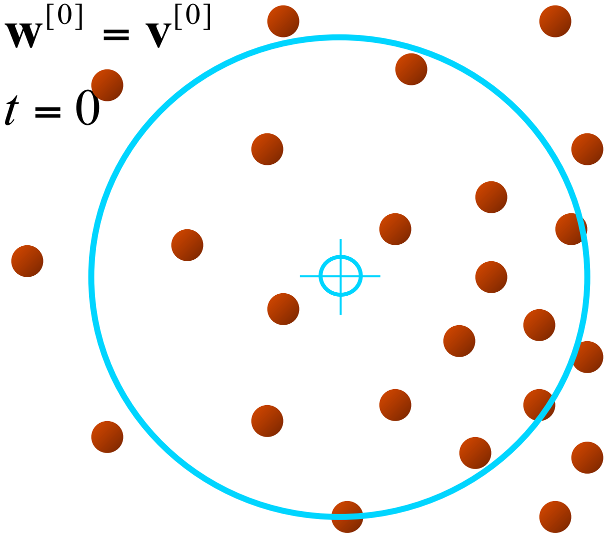
- 2. Initialize

$$\mathbf{H} = \begin{bmatrix} h_s^2 & 0 & 0 & 0 & 0 \\ 0 & h_s^2 & 0 & 0 & 0 \\ 0 & 0 & h_{red}^2 & 0 & 0 \\ 0 & 0 & 0 & h_{green}^2 & 0 \\ 0 & 0 & 0 & 0 & h_{blue}^2 \end{bmatrix}$$

and

$$\mathbf{w}^{[0]} = \mathbf{v}^{[0]}$$

$$t = 0$$



h_s spatial scale

h_r, h_g, h_b color scales

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

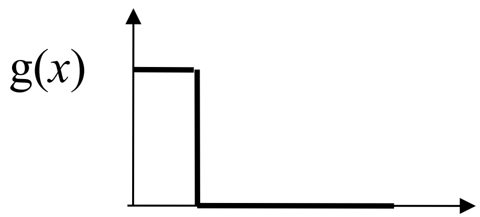
Mean-Shift Filtering Procedures

[Comaniciu 99']

- 3. Compute the centre of mass **w**

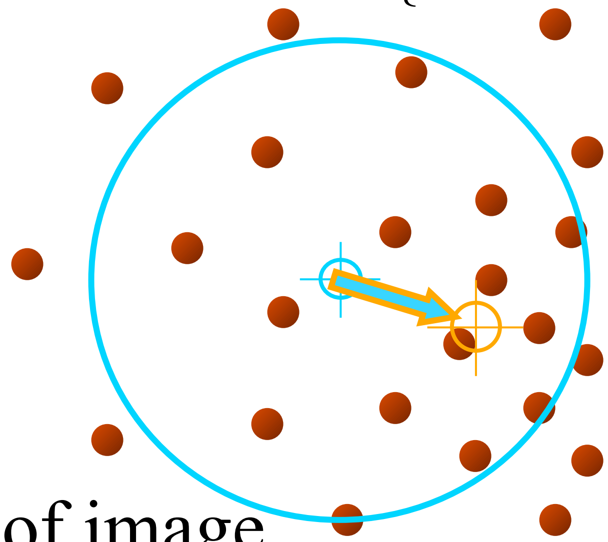
$$\mathbf{w}^{[t+1]} = \frac{\sum_{i=1}^{n} g\left(d^2\left(\mathbf{w}^{[t]}, \mathbf{v}_i, \mathbf{H}\right)\right) v_i}{\sum_{i=1}^{n} g\left(d^2\left(\mathbf{w}^{[t]}, \mathbf{v}_i, \mathbf{H}\right)\right)}$$

d(.): Mahalanobis distance



$$g(x) = \begin{cases} 1 & 0 < x \leq 1 \\ 0 & otherwise \end{cases}$$

until convergence



- 4. Make $v^{[t]} = w^{[\text{mode}]}$
- 5. Repeat steps 2-4 for each pixel of image

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Our Problem: optimize bandwidth H

- Two questions:
  - 1. Can we use gradient ascent to optimize **H**?
  - 2. Is it worth using an anisotropic bandwidth **H**?

$$\mathbf{H} = \begin{bmatrix} h_s^2 & 0 & 0 & 0 & 0 \\ 0 & h_s^2 & 0 & 0 & 0 \\ 0 & 0 & h_r^2 & 0 & 0 \\ 0 & 0 & 0 & h_r^2 & 0 \\ 0 & 0 & 0 & 0 & h_r^2 \end{bmatrix}$$

Isotropic H

$$\mathbf{H} = \begin{bmatrix} h_s^2 & 0 & 0 & 0 & 0 \\ 0 & h_s^2 & 0 & 0 & 0 \\ 0 & 0 & h_{red}^2 & 0 & 0 \\ 0 & 0 & 0 & h_{green}^2 & 0 \\ 0 & 0 & 0 & 0 & h_{blue}^2 \end{bmatrix}$$

Anisotropic H

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Description of Experiments

- Tested image

Original Image

Noisy Image

$\sigma^2 = 30$

- Isotropic Bandwidth **H**

$$\mathbf{H} = \begin{bmatrix} h_s^2 & 0 & 0 & 0 & 0 \\ 0 & h_s^2 & 0 & 0 & 0 \\ 0 & 0 & h_r^2 & 0 & 0 \\ 0 & 0 & 0 & h_r^2 & 0 \\ 0 & 0 & 0 & 0 & h_r^2 \end{bmatrix} \quad \begin{matrix} h_s \in [1,5] \\ h_r \in [20,240] \end{matrix}$$

Anisotropic Bandwidth **H**

$$\mathbf{H} = \begin{bmatrix} h_s^2 & 0 & 0 & 0 & 0 \\ 0 & h_s^2 & 0 & 0 & 0 \\ 0 & 0 & h_{red}^2 & 0 & 0 \\ 0 & 0 & 0 & h_{green}^2 & 0 \\ 0 & 0 & 0 & 0 & h_{blue}^2 \end{bmatrix} \quad \begin{matrix} h_s \in [1,5] \\ h_{red} \in [20,240] \\ h_{green} \in [20,240] \\ h_{blue} \in [20,240] \end{matrix}$$

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Results isotropic H VS anisotropic H

* Results Validation: PSNR

- PSNR results for anisotropic **H** is a little better

(a) Original Image

$\sigma^2 = 30$

(b) Noisy Image
19.71dB(ref)

(c) Optimal for
isotropic H
28.93dB(3,160)

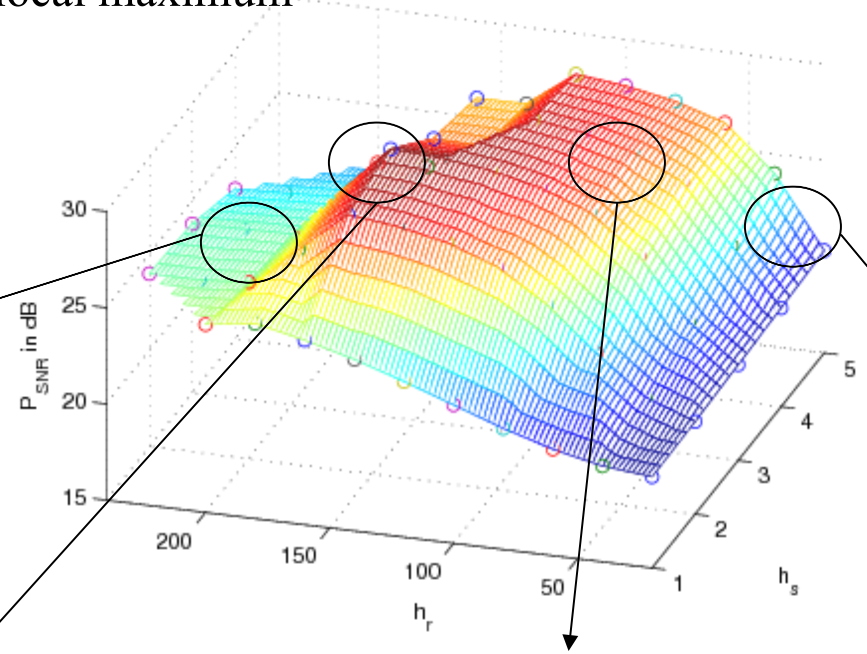
(d) Optimal for
anisotropic parameters
29.01dB
(3,240,140,200)

- Gradient Ascent (GA) can optimize bandwidth **H**
  - For isotropic **H**, GA is robust for searching the optimization.
  - For anisotropic **H**, GA can be used if the neighborhood radius is > 2.5

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Results (PSNR evolution example against hs&hr of isotropic H)

- Gradient Ascent algorithm
  - PSNR has no local maximum



24.17dB(2,200)

20.32dB(5,20)

28.93dB(3,160)

27.96dB(4,100)

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Perspectives

- Mean-Shift filtering experiments on more data
  - Synthesized data & RGB images & Medical data

- Choice of bandwidth **H**

$$\mathbf{H}=\begin{bmatrix} h_s^2 & 0 & 0 & 0 & 0 \\ 0 & h_s^2 & 0 & 0 & 0 \\ 0 & 0 & h_r^2 & 0 & 0 \\ 0 & 0 & 0 & h_r^2 & 0 \\ 0 & 0 & 0 & 0 & h_r^2 \end{bmatrix}$$

Isotropic **H**

$$H=\begin{bmatrix} h_s^2 & 0 & 0 & 0 & 0 \\ 0 & h_s^2 & 0 & 0 & 0 \\ 0 & 0 & h_{red}^2 & 0 & 0 \\ 0 & 0 & 0 & h_{green}^2 & 0 \\ 0 & 0 & 0 & 0 & h_{blue}^2 \end{bmatrix}$$

Anisotropic **H**

Move here

$$H=\begin{bmatrix} h_s^2 & 0 & 0 & 0 & 0 \\ 0 & h_s^2 & 0 & 0 & 0 \\ 0 & 0 & h_{red}^2 & h_{rg}^2 & h_{rb}^2 \\ 0 & 0 & h_{rg}^2 & h_{green}^2 & h_{gb}^2 \\ 0 & 0 & h_{rb}^2 & h_{gb}^2 & h_{blue}^2 \end{bmatrix}$$

Range Correlated **H**

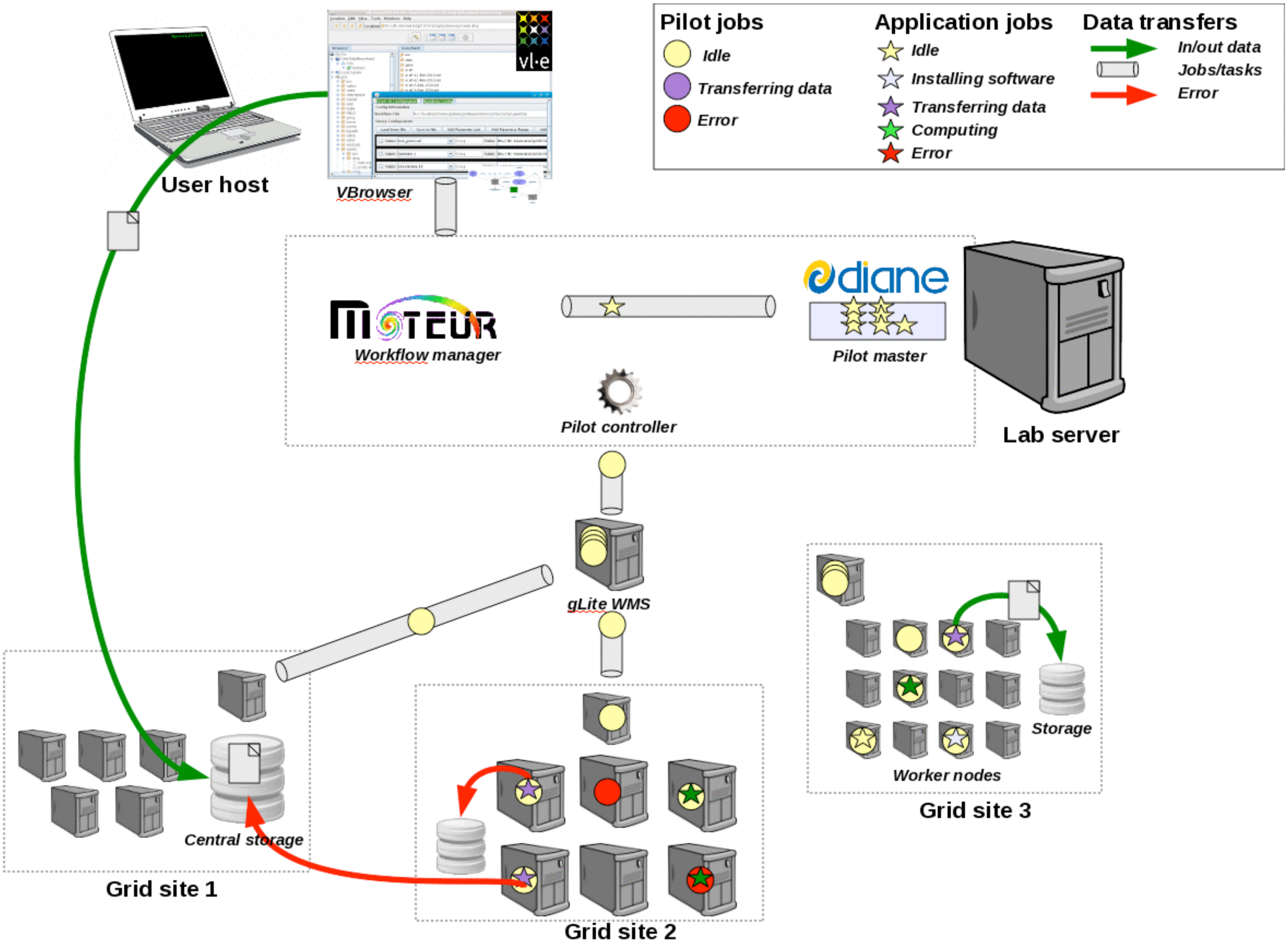
164 CPU days

19000 jobs

~100 GB

- Need for grid !

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

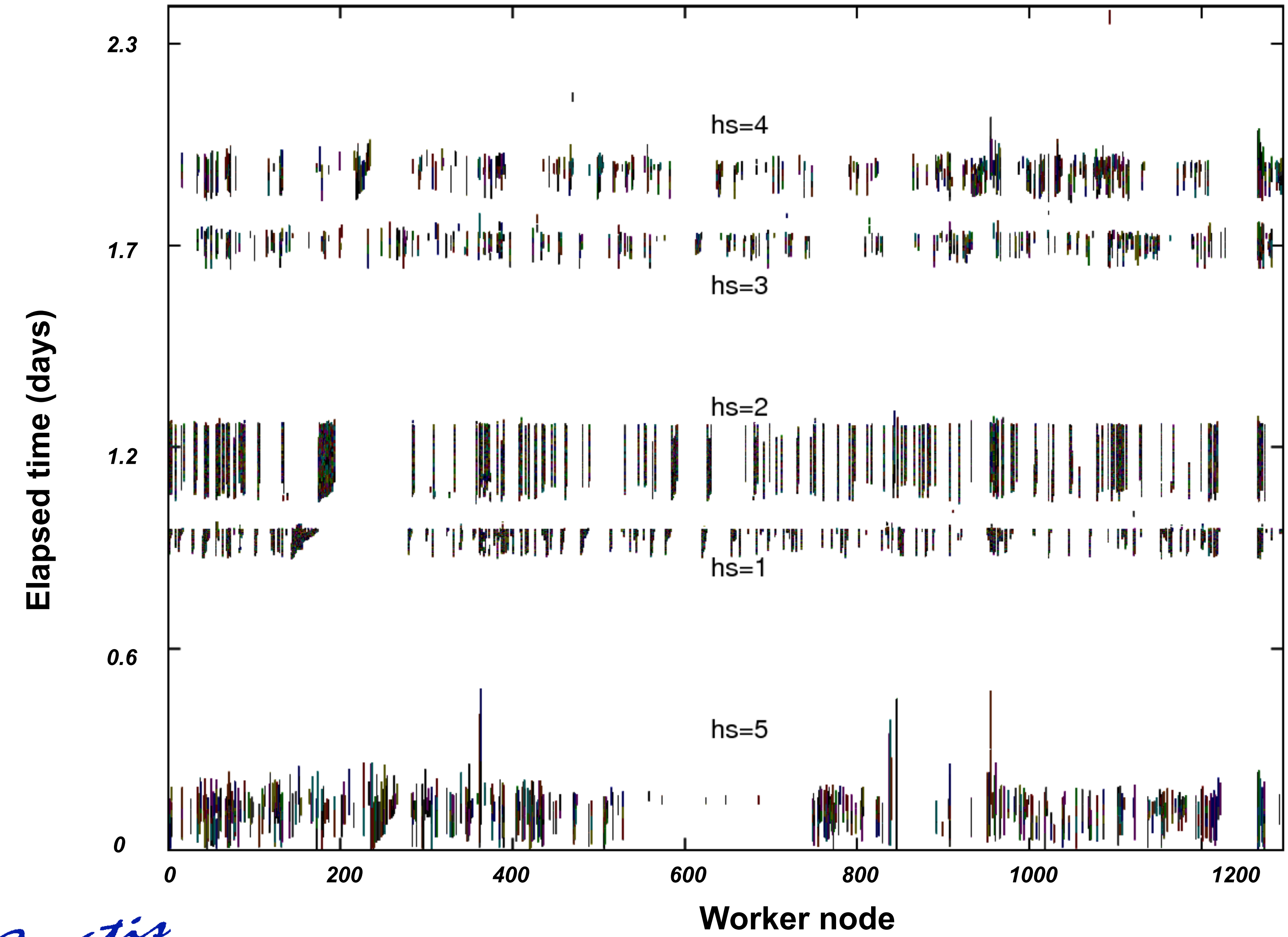

Pilot jobs
Idle
Transferring data
Error
Application jobs
Idle
Installing software
Transferring data
Computing
Error
Data transfers
In/out data
Jobs/tasks
Error
User host
VBrowser
vl·e
diane
Workflow manager
Pilot master
Pilot controller
Lab server
gLite WMS
Central storage
Grid site 1
Grid site 2
Storage
Worker nodes
Grid site 3

Grid Setup

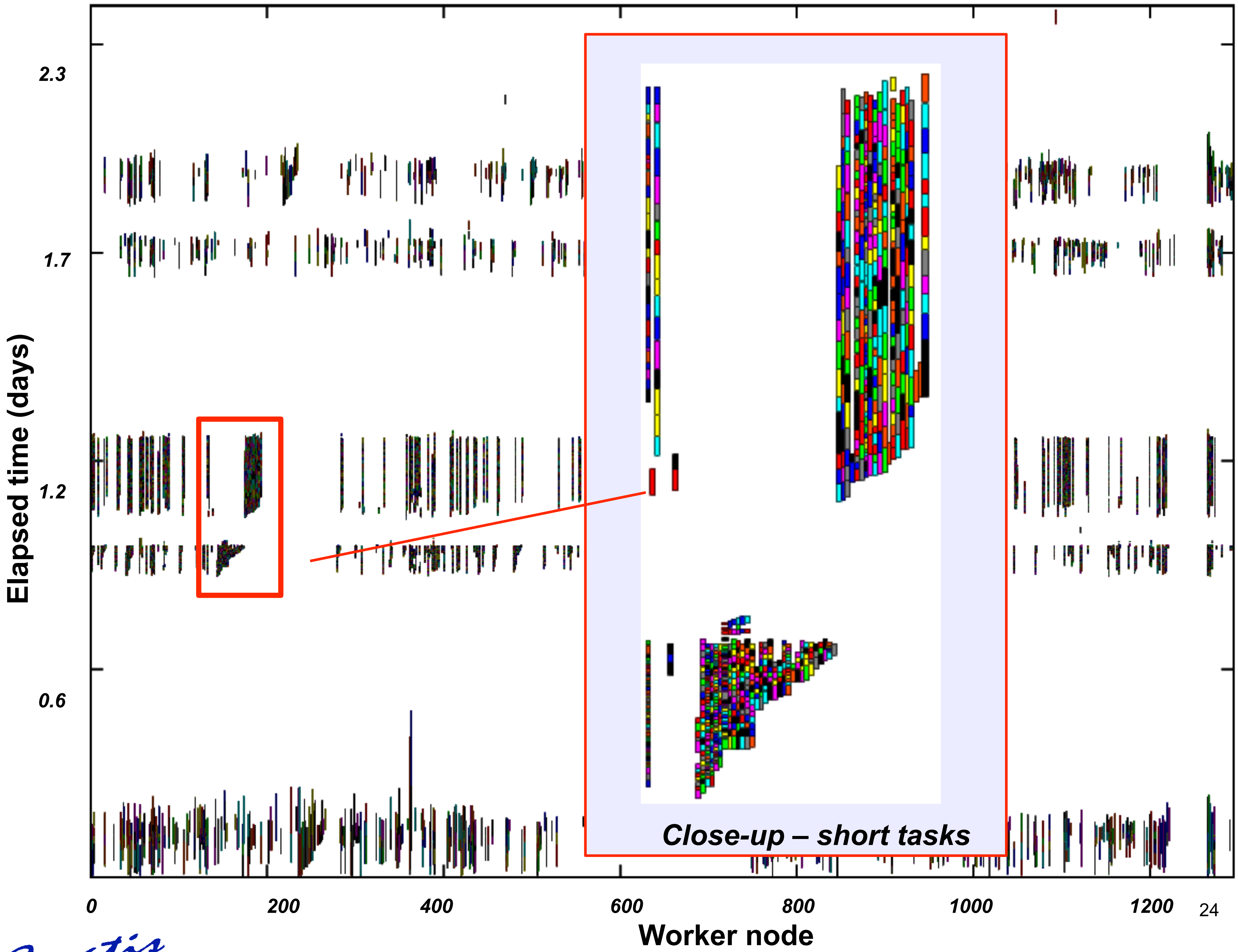
- Application porting
  - Compiled Matlab for Centos5, 64bits
  - Application described as a workflow
- Infrastructure
  - EGEE, biomed Virtual Organisation
  - ~120 sites
- Handling long jobs
  - Proxy uploaded to myproxy
  - Jobs have watchdog scripts to fetch fresh proxy
- Data transfer reliability
  - Data cached in pilots
  - Results written where produced (fall-back on central)
  - Upload test before execution

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

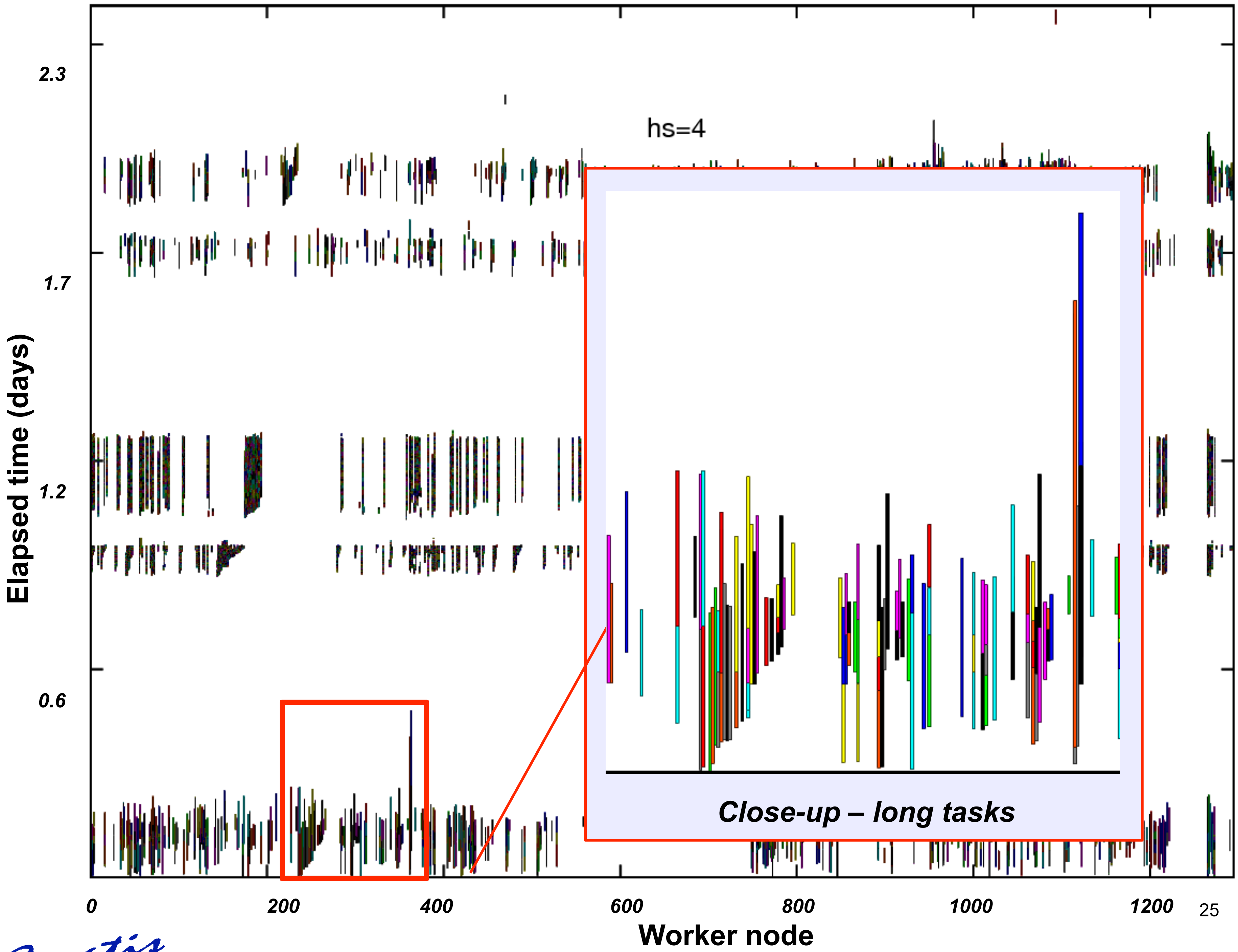
Gantt diagram of the experiment



Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr



Creatis
tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr



Creatis
tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Performance analysis

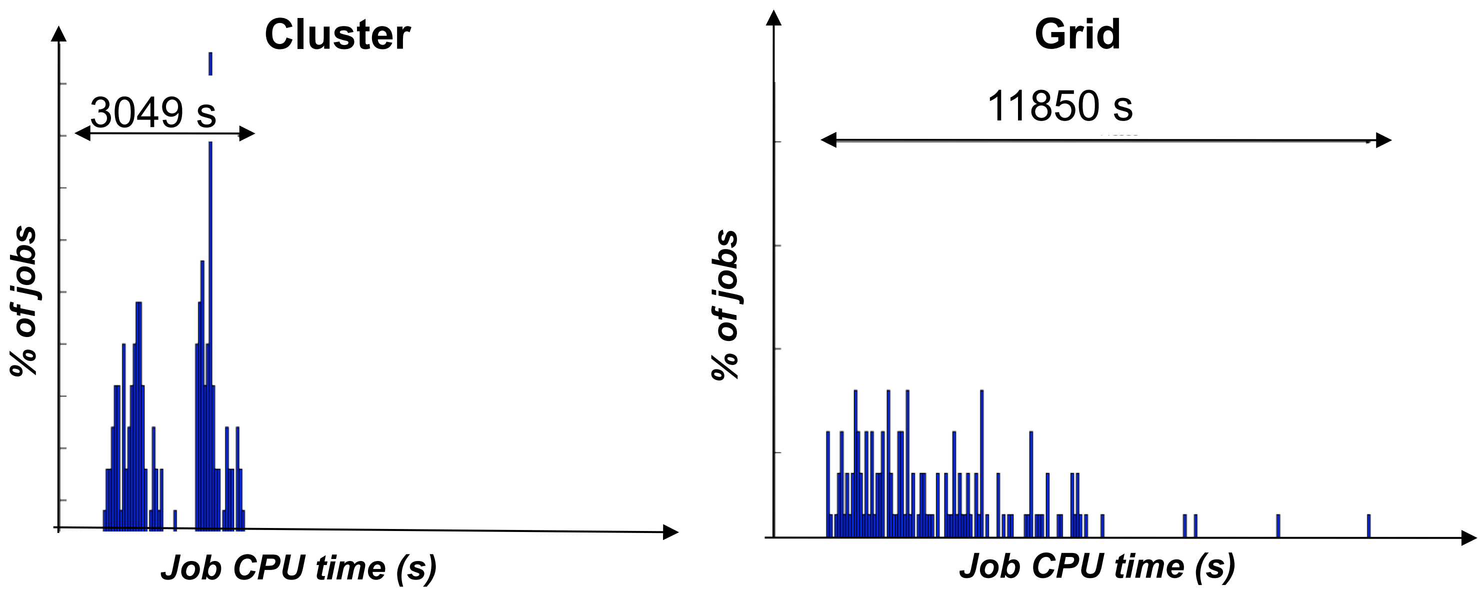
h_s value	Total CPU time	Elapsed time	Speed-up	Total data transfer time	Produced data	Successful tasks	Total tasks	Error ratio
1	13.0 days	3h25min	92	89.8h	43GB	8,000	8,106	1.3%
2	50.4 days	18h36min	65	63.2h	41GB	8,000	8,929	10.4%
3	17.3 days	13h01min	32	34.8h	5GB	1,000	1,317	24%
4	29.0 days	13h23min	52	44.7h	5GB	1,000	1,089	8.2%
5	54.3 days	11h09min	117	51.9h	5.1GB	1,000	1,179	15.2%

- Error causes
  - Cannot download input (76.7%) ; cannot upload output (9.9%)
  - Pilot/master communication problem (8.8%)
  - Application error (4.8%)
- Time facts
  - Total CPU time: 5.8 months
  - Average IO / CPU: 7%

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Comparison with cluster execution

- Local cluster
  - Six 8-core Intel Xeon L5430 2.66 Ghz with PBS
  - (average bogomips cluster) / (average bogomips grid) = 1.14
- CPU time comparison
  - (Total CPU time grid) / (Total CPU time cluster) = 1.5
  - Increased variability on grid (factor 4):



tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Conclusion (on Mean-Shift side)

- Mean-Shift results
  - Worth using bandwidth with 4 parameters (anisotropic bandwidth **H**)
  - Optimization with gradient ascent is possible

- Future work on Mean-shift
  - Validate Mean-Shift results on more images (including medical images)
  - Test bandwidth with 7 parameters
  - Try other quality measurement metrics (besides PSNR)

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Conclusion (on grid side)

- Grid achievements
  - Significant speed-up (164 days => 2.5 days)
  - Autonomous usage from end-user
  - Some observations on scheduling and data transfers (see paper)

- Future work on the grid setup
  - Improve experiment planning on grids
  - Make data retrieval more robust

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Credits

EGEE-III Life-Science cluster

Institut des Grilles du CNRS

Creatis

Fabrice Bellet

MOTEUR

Johan Montagnat, Alban Gaignard ; CNRS I3S

VL-e medical software

Silvia D. Olabarriaga ; *AMC Amsterdam*

Piter T. de Boer ; Universiteit Van Amsterdam

Spiros Koulouzis ; Universiteit Van Amsterdam

Pilot jobs (DIANE)

Jakub T. Moscicki ; *CERN*

Grid support

China Scholarship Counsel (CSC)

https://gus.fzk.de

tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr

Thank You for your attention!

Creatis tristan.glatard@creatis.insa-lyon.fr ; ting.li@creatis.insa-lyon.fr