La réalité augmentée

5. POSITIONNEMENT GLOBAL

5.1 Le descripteur SIFT et les BoW
5.2 Les descripteurs CNN (*Convnet features*)
5.3 Localisation basée objets

5. Positionnement global

5.1 LE DESCRIPTEUR SIFT ET LES BOW

Le descripteur SIFT

- SIFT = *Scale Invariant Feature Transform*
- SIFT est à la fois :
  - Un détecteurs de points-clés (*blobs*) : centres de "tâches" ou "gouttes" détectées dans l'image,
  - Un descripteur de points : à chaque blob est associé un vecteur de taille 128, invariant par changements d'échelle et rotations dans l'image
- De multiples variantes existent : par exemple, garder le descripteur mais utiliser un détecteur plus rapide pour permettre l'utilisation sur téléphone mobile

Le descripteur SIFT

- Principe du détecteur : l'image f(x,y) est convoluée avec un noyau gaussien g(x,y,t) sur un continuum d'échelles t ≥ 0

$$L(x,y;t) = g(x,y,t) * f(x,y)$$

$$g(x,y,t) = \frac{1}{2\pi t^2} e^{-(x^2+y^2)/(2t^2)}$$

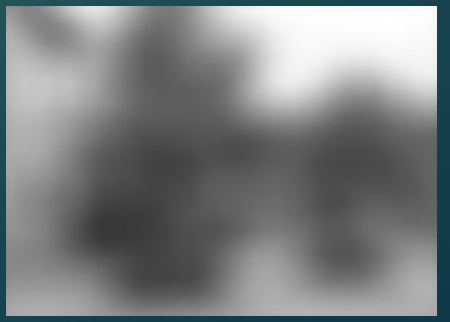
t=0

t=1

t=4

t=16

t=32



t=64

Le descripteur SIFT

- L'image originale peut être vue comme une plaque diffusant la chaleur, avec des zones froides et des zones chaudes, suivant les niveaux de gris

Le descripteur SIFT

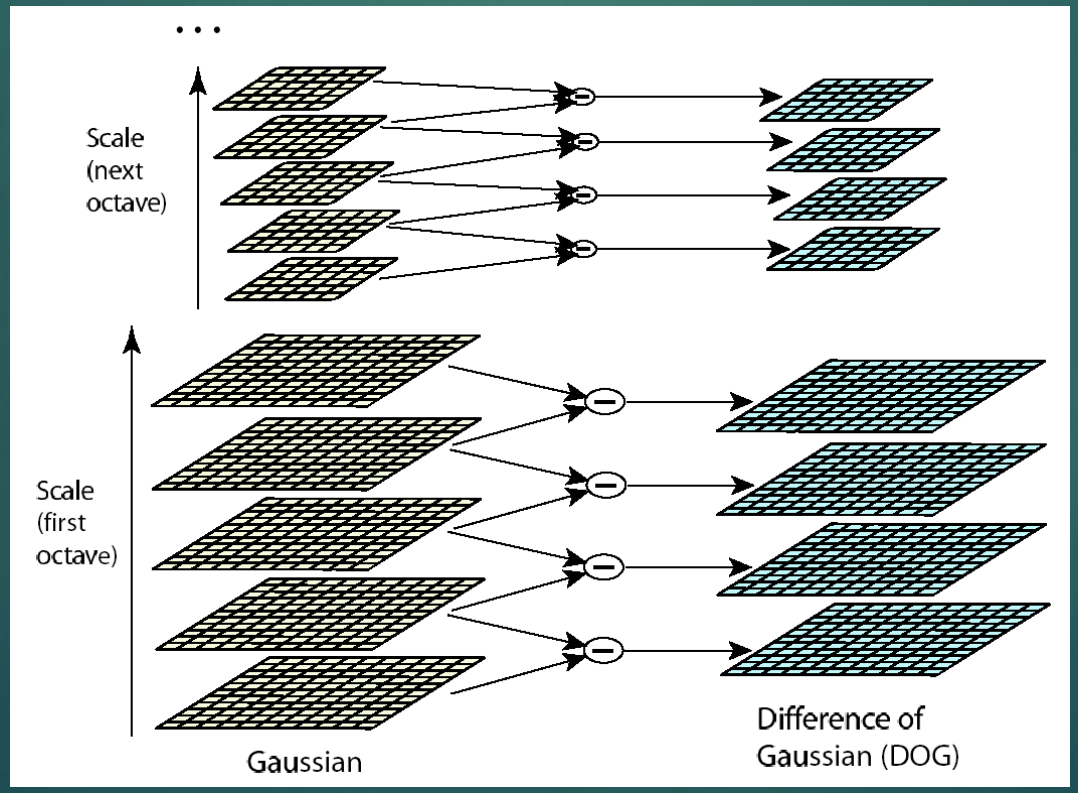
- Equation de la chaleur

$$\partial_t L = \frac{1}{2}\nabla^2 L$$

- L'étalement spatial des blobs (laplacien) peut être mesuré localement en considérant l'evolution temporelle des niveaux de gris des points situés au centre des blobs
- Différence entre deux instants = différence de gaussiennes (DoG)
- Une DoG est plus rapide à calculer qu'un laplacien

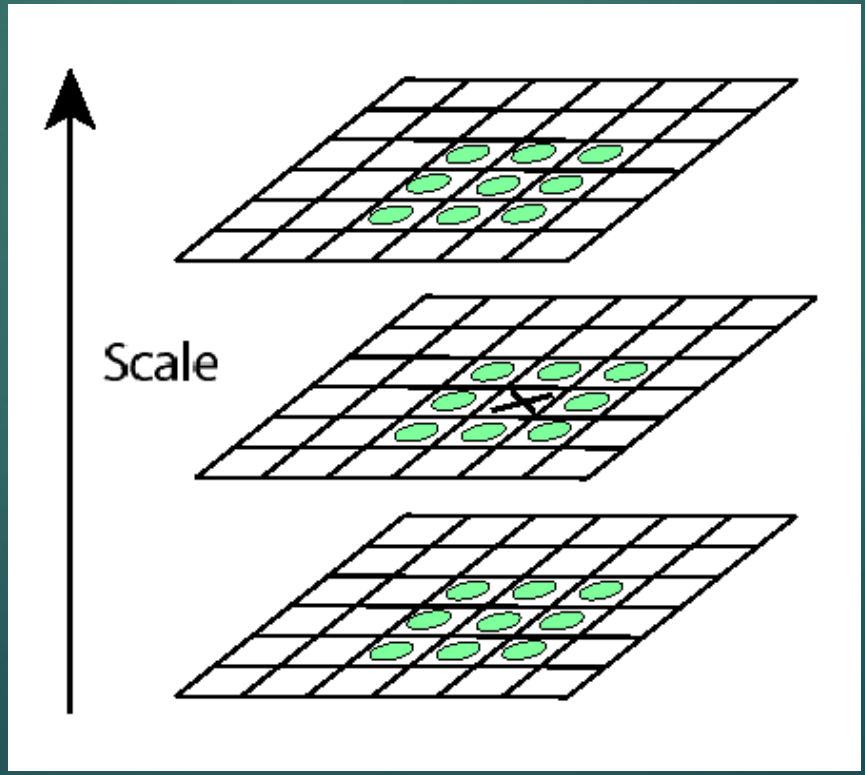
Le descripteur SIFT

- Utilisation d'une pyramide de DoG (espace des échelles)



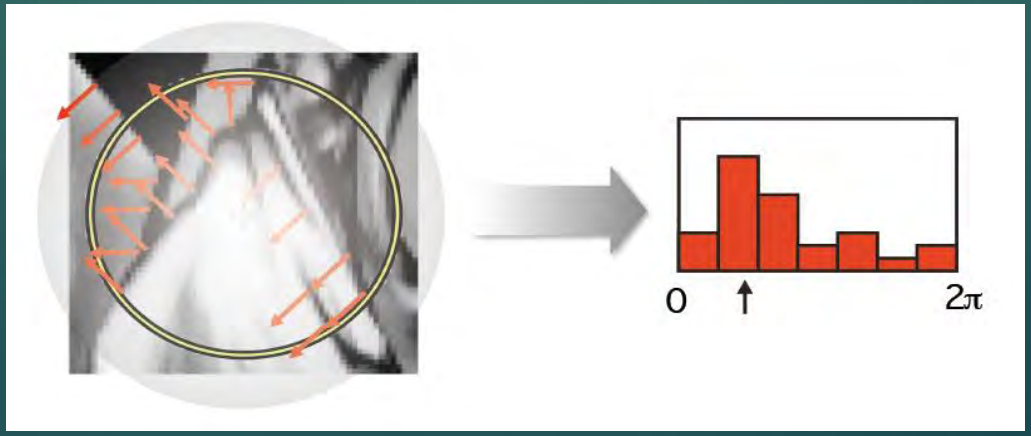
Le descripteur SIFT

- Localisation des points-clé : extrema des DoG dans l'espace des échelles
- Point important : une échelle est associée à chaque point-clé



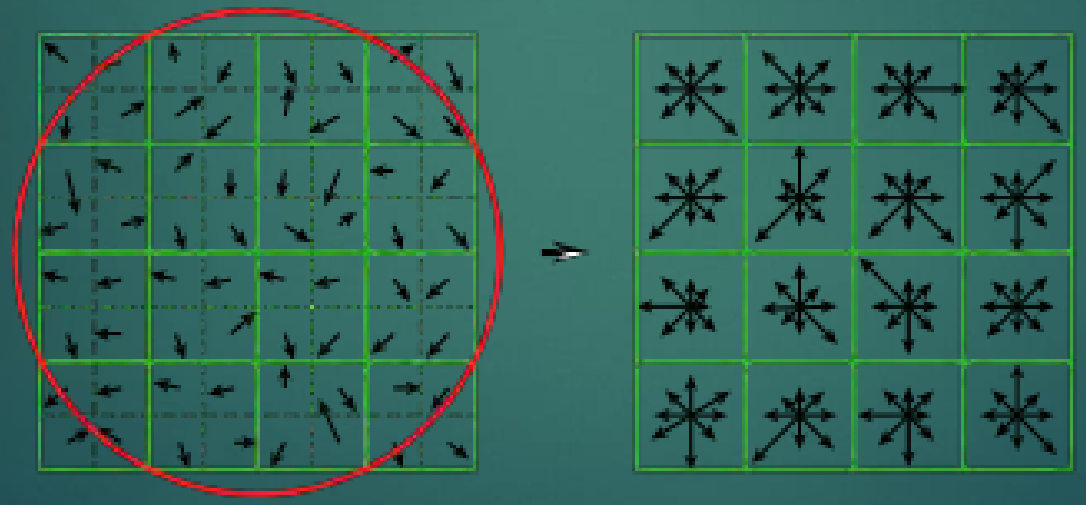
Le descripteur SIFT

- Descripteur : histogramme des orientations des gradients de l'image autour du point-clé
- Calculé à la résolution correspondant à l'échelle du point-clé
  → invariance aux changements d'échelle
- Les histogrammes sont décalés circulairement de manière à ramener le plus grand pic au début de l'histogramme
  → invariance à la rotation

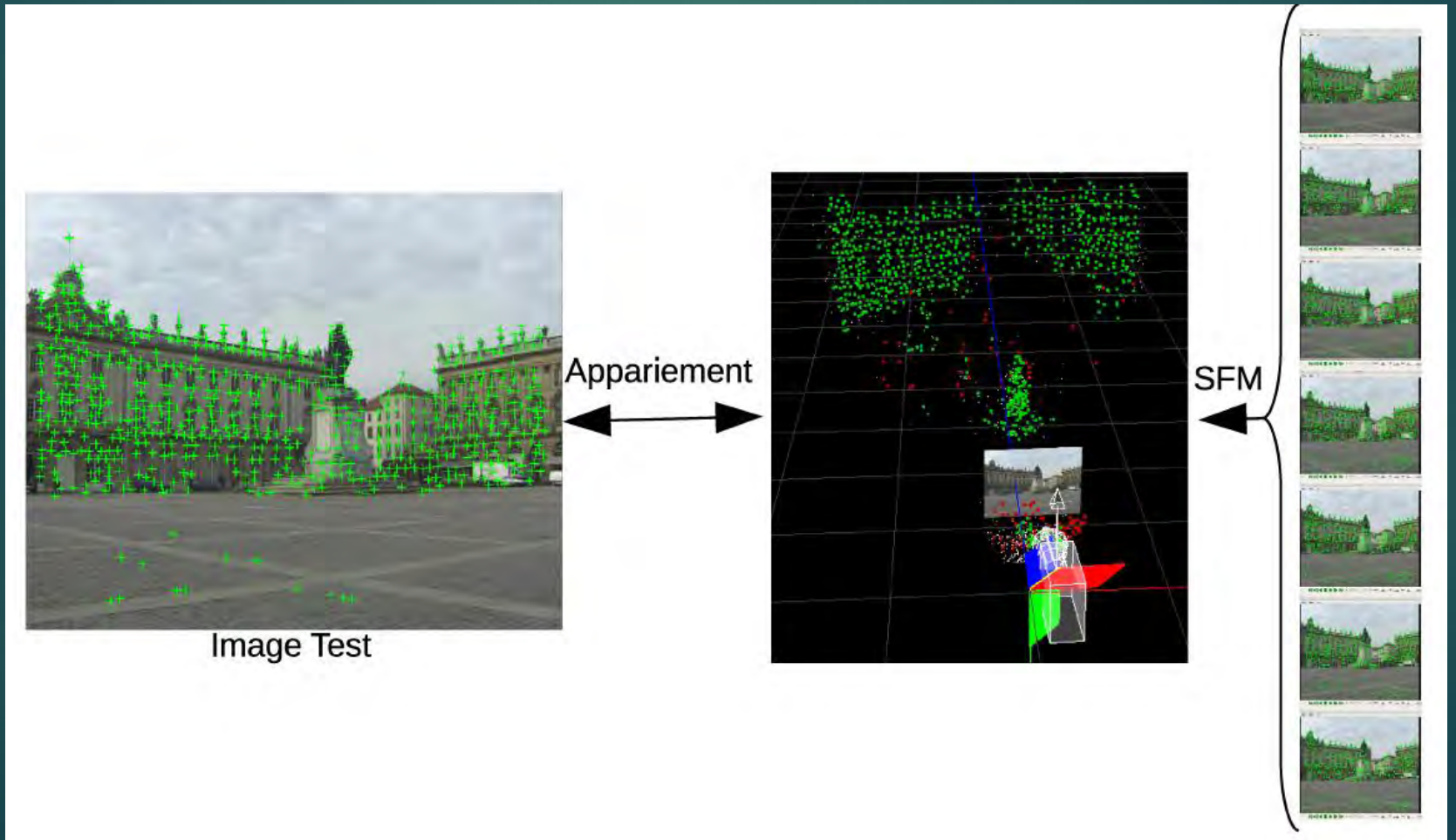


Le descripteur SIFT

- En pratique : 8 orientations x 4 x 4 histograms = vecteur de taille 128
- Appariement : distance euclidienne entre les vecteurs de l'image courante et les vecteurs de la base (kd-tree)

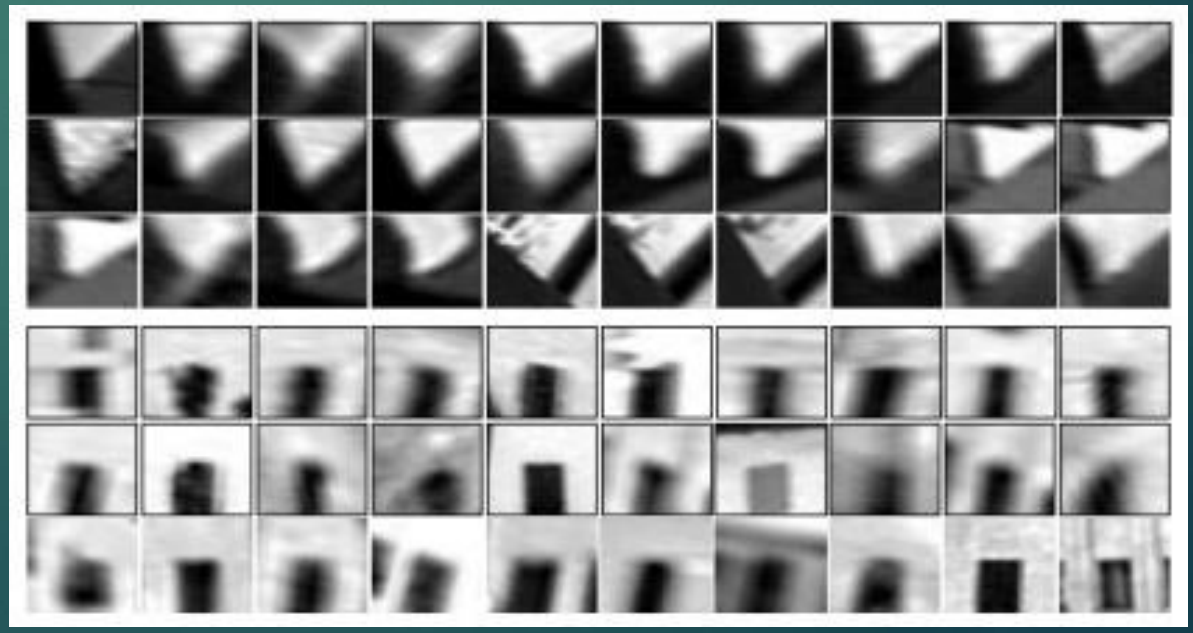


Positionnement par appariement de descripteurs SIFT



Problème des grands environnements

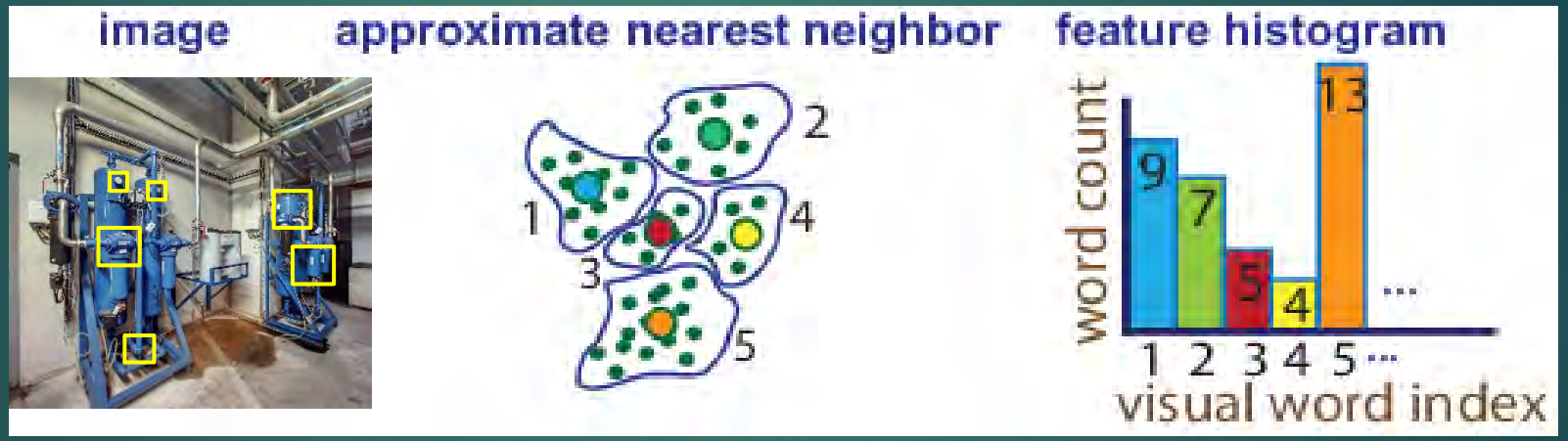
- Dans les grands environnements, le nombre d'indices visuels à parcourir peut être gigantesque
- Idée pour accélérer l'appariement : regrouper les descripteurs en K clusters
  - Chaque cluster est appelé « mot visuel » (visual word)
  - L'ensemble des mots visuels forme un vocabulaire appelé « sac de mots » (bag-of-words – BoW)



© « Video google »
VGG Oxford

BoW : application à la reconnaissance de scènes

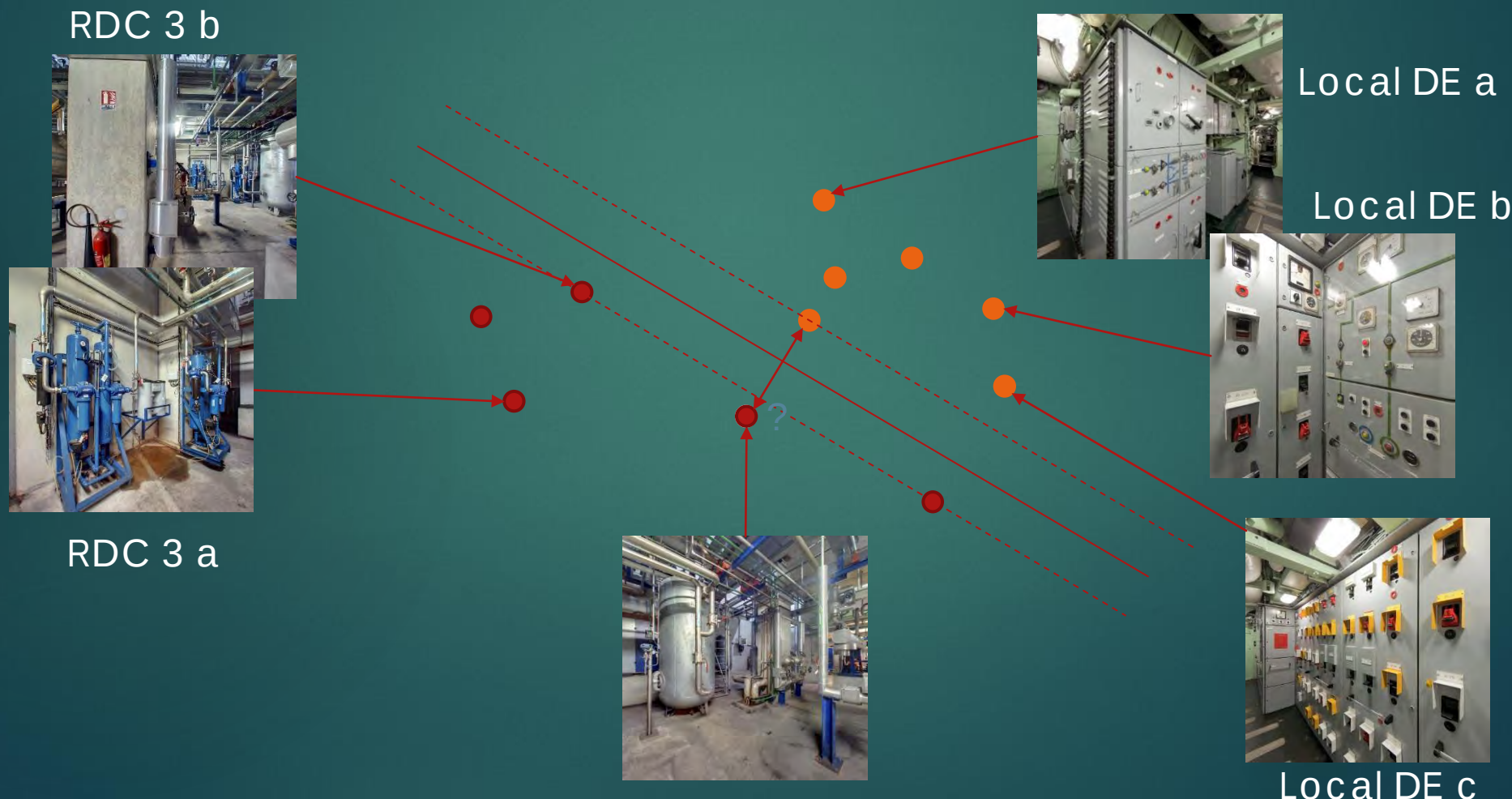
- Chaque scène est représentée par un ensemble d'images
- Chaque image de cette base est décrite par l'histogramme des mots visuels apparaissant dans l'image



- L'image de la scène à reconnaître est elle-même décrite par son histogramme de mots visuels

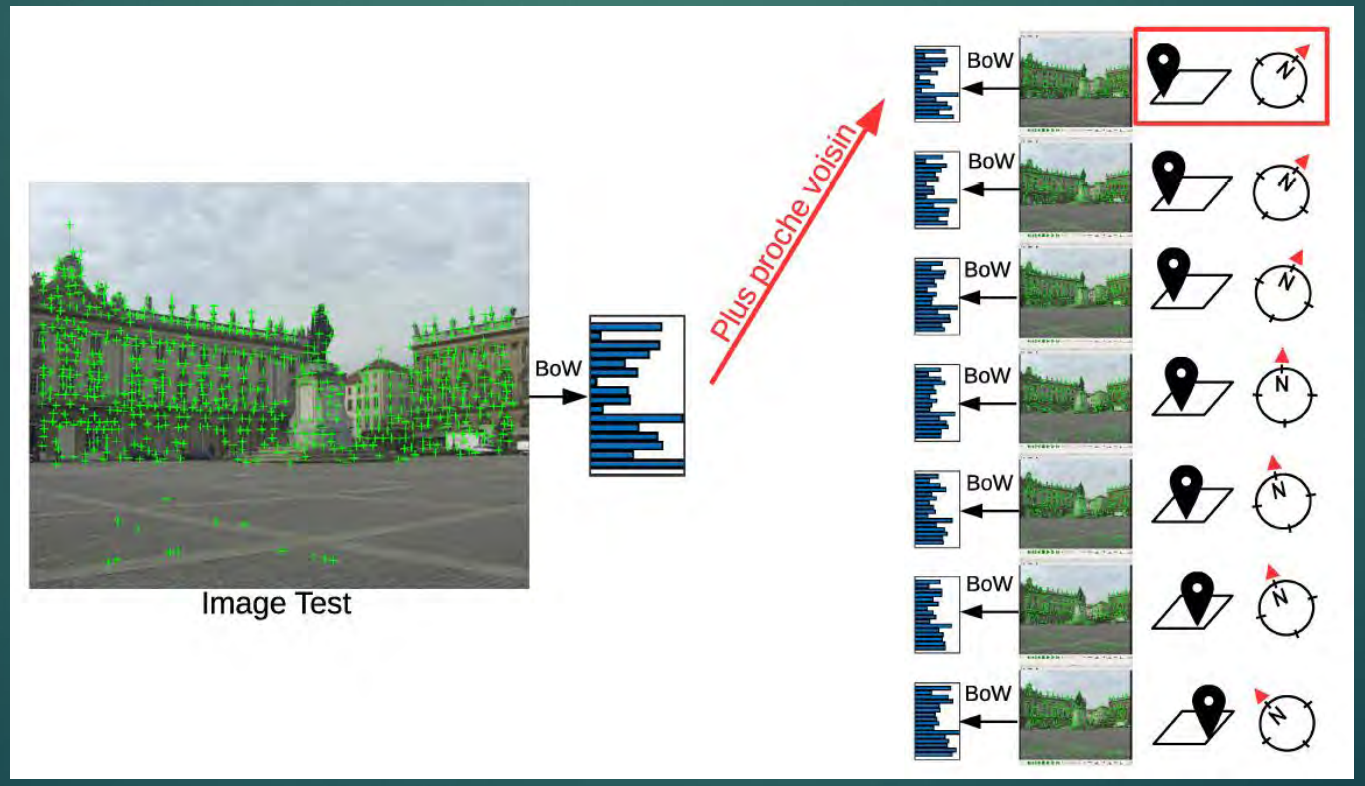
BoW : application à la reconnaissance de scènes

- Classification par machine à vecteurs de support (SVM)



BoW : application au positionnement par plus proche voisin (descripteurs globaux)

- Comparaison d'histogrammes de mots visuels entre une image test et des images de référence associées à des poses



- Robuste mais peu précis + requiert de nombreuses images ground truth

Facteurs pénalisant l'appariement de mots visuels

- Grands changements de perspective

Facteurs pénalisant l'appariement de mots visuels

- Lumière, météo, ombrages

Facteurs pénalisant l'appariement de mots visuels

- Alternance jour/nuit

Facteurs pénalisant l'appariement de mots visuels

- Résolution et flou optique

Facteurs pénalisant l'appariement de mots visuels

- Artéfacts de non-planarité et occultations

Descripteurs globaux versus appariement de mots visuels

- Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions [Sattler et al., 2018]

Descripteurs globaux versus appariement de mots visuels

Appariement de mots visuels

Descripteurs globaux

	Aachen				CMU								
	day		night		foliage		mixed foliage		no foliage		urban	suburban	park
m	.25/.50/5.0		0.5/1.0/5.0		.25/.50/5.0		.25/.50/5.0		.25/.50/5.0		.25/.50/5.0	.25/.50/5.0	.25/.50/5.0
deg	2/5/10		2/5/10		2/5/10		2/5/10		2/5/10		2/5/10	2/5/10	2/5/10
Active Search	57.3 / 83.7 / 96.6		19.4 / 30.6 / 43.9		28.8 / 32.5 / 35.9		25.1 / 29.4 / 33.9		52.5 / 59.4 / 66.7		55.2 / 60.3 / 65.1	20.7 / 25.9 / 29.9	12.7 / 16.3 / 20.8
CSL	52.3 / 80.0 / 94.3		24.5 / 33.7 / 49.0		16.3 / 19.1 / 26.0		15.2 / 18.8 / 28.6		36.5 / 43.2 / 57.5		36.7 / 42.0 / 53.1	8.6 / 11.7 / 21.1	7.0 / 9.6 / 17.0
DenseVLAD	0.0 / 0.1 / 22.8		0.0 / 2.0 / 14.3		13.2 / 31.6 / 82.3		16.2 / 38.1 / 85.4		17.8 / 42.1 / 91.3		22.2 / 48.7 / 92.8	9.9 / 26.6 / 85.2	10.3 / 27.0 / 77.0
NetVLAD	0.0 / 0.2 / 18.9		0.0 / 2.0 / 12.2		10.4 / 26.1 / 80.1		11.0 / 26.7 / 78.4		11.8 / 29.1 / 82.0		17.4 / 40.3 / 93.2	7.7 / 21.0 / 80.5	5.6 / 15.7 / 65.8
FABMAP	0.0 / 0.0 / 4.6		0.0 / 0.0 / 0.0		1.1 / 2.7 / 16.5		1.0 / 2.5 / 14.7		3.6 / 7.9 / 30.7		2.7 / 6.4 / 27.3	0.5 / 1.5 / 13.6	0.8 / 1.7 / 11.5

	day conditions							night conditions	
	dawn	dusk	OC-summer	OC-winter	rain	snow	sun	night	night-rain
m	.25 / .50 / 5.0	.25 / .50 / 5.0	.25 / .50 / 5.0	.25 / .50 / 5.0	.25 / .50 / 5.0	.25 / .50 / 5.0	.25 / .50 / 5.0	.25 / .50 / 5.0	.25 / .50 / 5.0
deg	2 / 5 / 10	2 / 5 / 10	2 / 5 / 10	2 / 5 / 10	2 / 5 / 10	2 / 5 / 10	2 / 5 / 10	2 / 5 / 10	2 / 5 / 10
ActiveSearch	36.2 / 68.9 / 89.4	44.7 / 74.6 / 95.9	24.8 / 63.9 / 95.5	33.1 / 71.5 / 93.8	51.3 / 79.8 / 96.9	36.6 / 72.2 / 93.7	25.0 / 46.5 / 69.1	0.5 / 1.1 / 3.4	1.4 / 3.0 / 5.2
CSL	47.2 /73.3 / 90.1	56.6 / 82.7 / 95.9	34.1 / 71.1 / 93.5	39.5 / 75.9 / 92.3	59.6 / 83.1 / 97.6	53.2 / 83.6 / 92.4	28.0 / 47.0 / 70.4	0.2 / 0.9 / 5.3	0.9 / 4.3 / 9.1
DenseVLAD	8.7 / 36.9 / 92.5	10.2 / 38.8 / 94.2	6.0 / 29.8 / 92.0	4.1 / 26.9 / 93.3	10.2 / 40.6 / 96.9	8.6 / 30.1 / 90.2	5.7 / 16.3 / 80.2	0.9 / 3.4 / 19.9	1.1 / 5.5 / 25.5
NetVLAD	6.2 / 22.8 / 82.6	7.4 / 29.7 / 92.9	6.5 / 29.6 / 95.2	2.8 / 26.2 / 92.6	9.0 / 35.9 / 96.0	7.0 / 25.2 / 91.8	5.7 / 16.5 / 86.7	0.2 / 1.8 / 15.5	0.5 / 2.7 / 16.4
FABMAP	1.2 / 5.6 / 14.9	4.1 / 18.3 / 55.1	0.9 / 8.9 / 39.3	2.6 / 13.3 / 44.1	8.8 / 32.1 / 86.5	2.0 / 8.2 / 28.4	0.0 / 0.0 / 2.4	0.0 / 0.0 / 0.0	0.0 / 0.0 / 0.0

Table 4. Evaluation on the **RobotCar Seasons** dataset. We report the percentage of queries localized within the three thresholds.