Imagerie et reconnaissance d'objet

N. Hascoët – Prof. F. Chinesta

5 juillet 2019

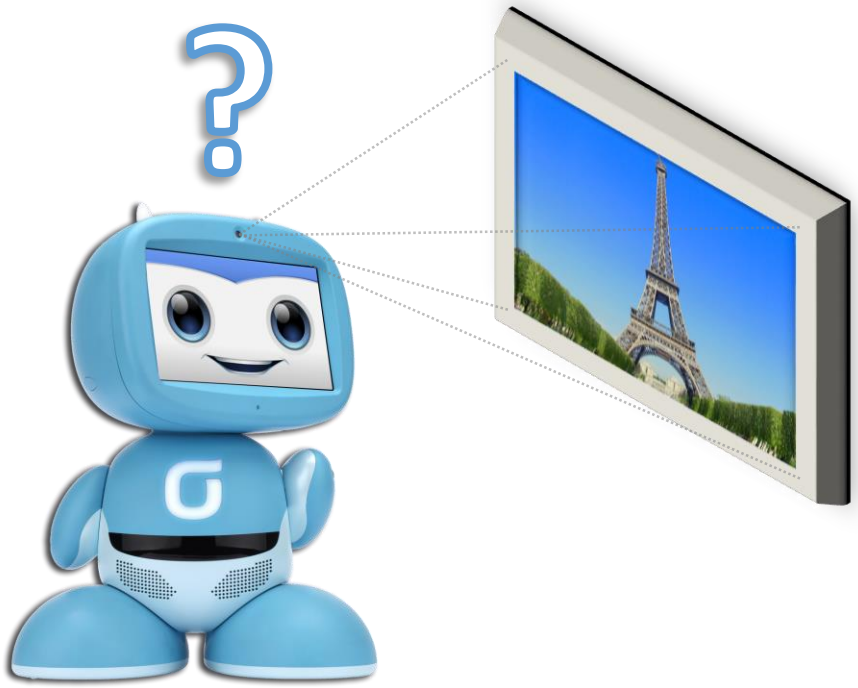
ARTS ET MÉTIERS ParisTech

Plan

- Introduction
- Etat de l’art et contributions
- Bases de données d’expérimentation
- Classification globale des descripteurs locaux
- Modèles SVM adaptés par classe de bâtiments
- Vérification et correction géométrique
- Recherche par similarité : résultats expérimentaux
- Conclusion

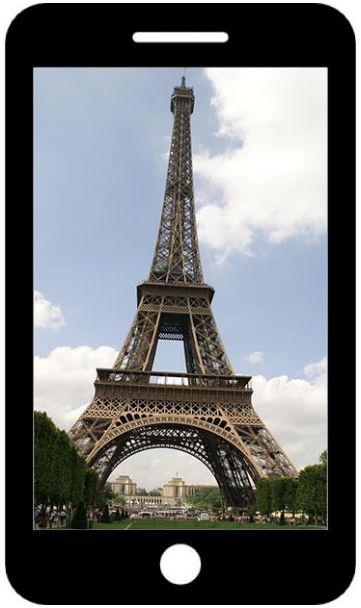
Introduction

- Vision par ordinateur : permet a un terminal de comprendre des images grâce à un système d'acquisition
- Reconnaissance faciale
- Interprétation de scènes urbaines

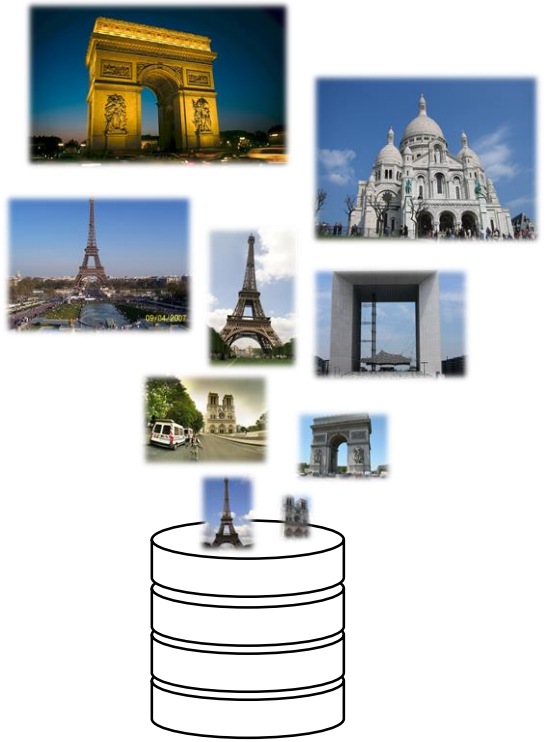


Introduction

- Objectif : reconnaître un bâtiment/monument d'intérêt de manière automatique à partir d'une photo



?



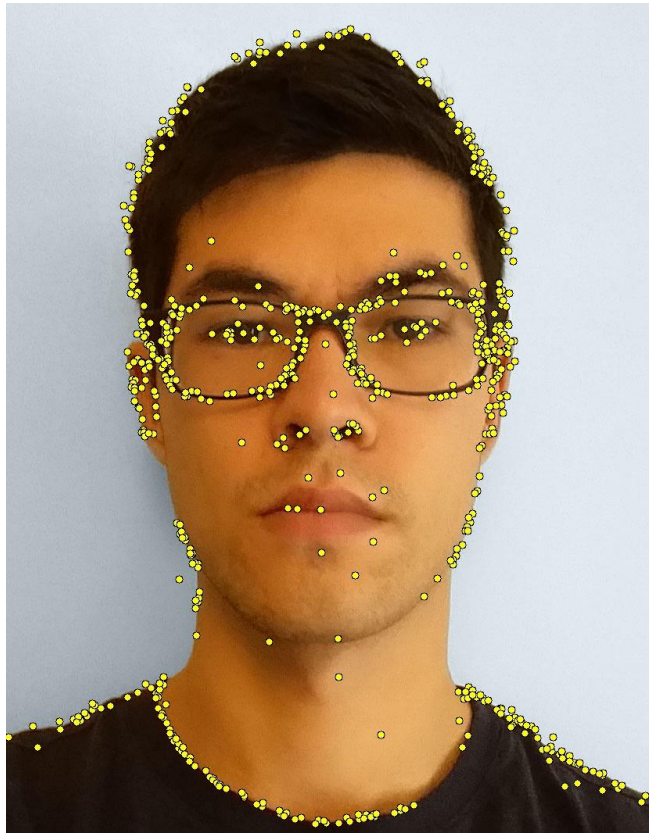
1. 
2. 
3. 
4. 
5. 

...

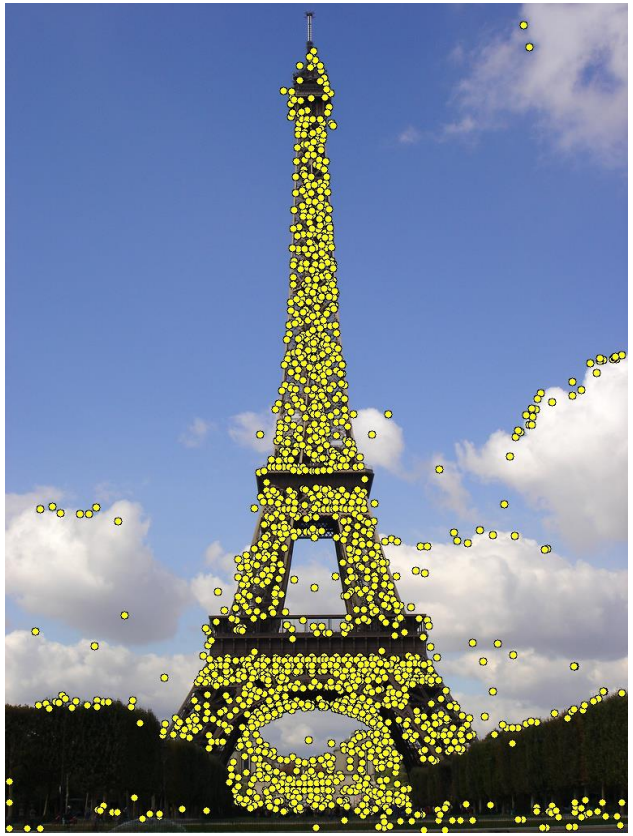
Introduction

- Défis et difficultés : une SURABONDANCE des points d'intérêt

689 points d'intérêt



2211 points d'intérêt



Introduction

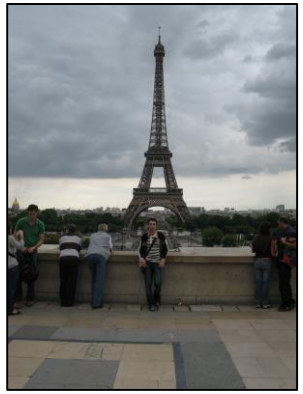
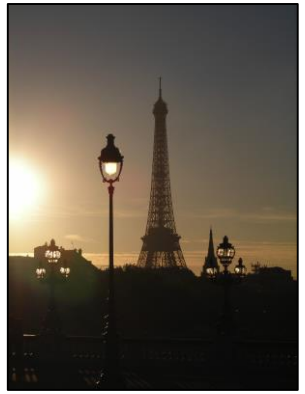
- Défis et difficultés : acquisitions selon de POINTS DE VUE différents





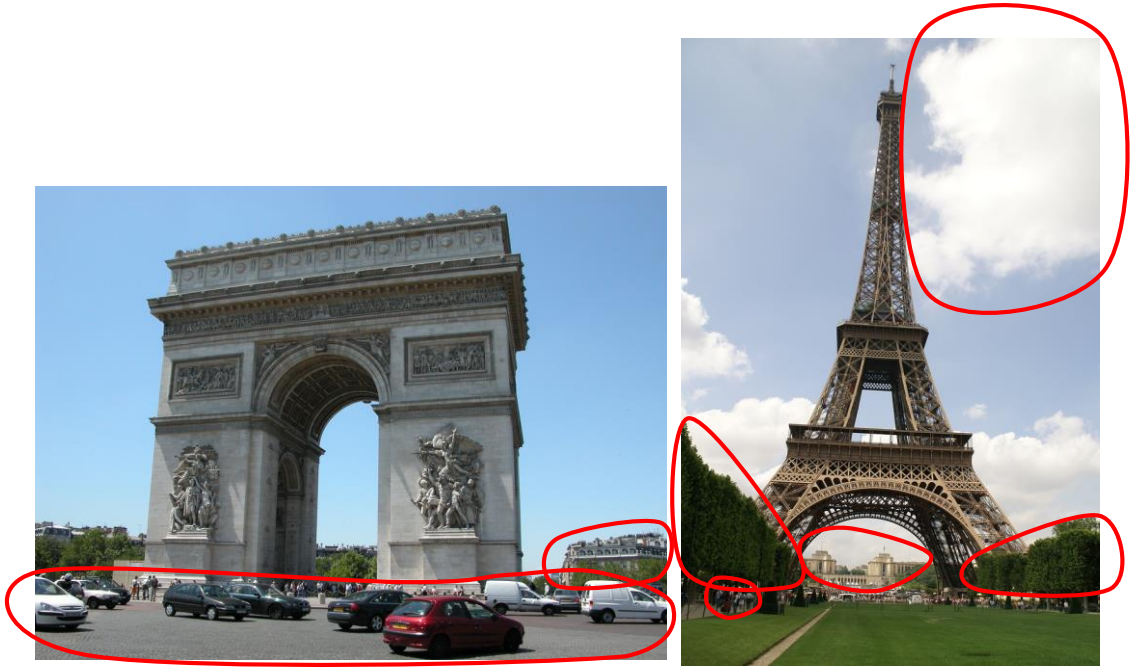
- Conditions d'ÉCLAIRAGE très variables




Introduction

- Défis et difficultés : COMPLEXITÉ des scènes URBAINES avec la présence de nombreux éléments « parasites »
  - Interférences avec
    - le ciel
    - les piétons
    - les véhicules
    - la végétation
    - d'autres bâtiments
    - etc.



Plan

- Introduction
- Etat de l’art et contributions
- Bases de données d’expérimentation
- Classification globale des descripteurs locaux
- Modèles SVM adaptés par classe de bâtiments
- Vérification et correction géométrique
- Recherche par similarité : résultats expérimentaux
- Conclusion

État de l'Art

- **Reconnaissance d'images pour la géolocalisation : descripteurs SIFT et représentation par BOW (Bag of Words)**
  - Gronat, Obozinski, Sivic, Pajdla : Learning and Calibrating Per-Location Classifiers for Visual Place Recognition ; International Journal of Computer Vision, 2013.
  - Philbin, Isard, Sivic, Zisserman : Descriptor Learning for Efficient Retrieval ; European Conference on Computer Vision, 2010.
  - Mikulik, Perdoch, Chum, Matas : Learning a Fine Vocabulary ; European Conference on Computer Vision, 2010.
  - Knopp, Sivic, Pajdla : Avoiding Confusing Features in Place Recognition ; European Conference on Computer Vision, 2010.
  - Schindler, Brown, Szeliski : City-Scale Location Recognition ; Computer Vision and Pattern Recognition, 2007.
- **Amélioration des performances du BOW en ajoutant une STRUCTURE d'images géo-marquées**
  - Sélection des points clefs
  - Reclassement des résultats

× Contraintes apportées par la reconnaissance de bâtiment en extérieur dans un environnement public

État de l'Art

- **Extension de requêtes**
  - Chum, Mikulik, Perdoch, Matas : Total recall II: Query Expansion Revisited ; Computer Vision and Pattern Recognition, 2011.
  - Philbin, Chum, Isard, Sivic, Zisserman : Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases ; Computer Vision and Pattern Recognition, 2008.
  - Arandjelovic, Zisserman : Three Things Everyone Should Know to Improve Object Retrieval ; Computer Vision and Pattern Recognition, 2012.
- **Échantillon requête extrait d'une image ⇒ Extension automatique de la région pour requête**

× Dépend de la sélection initiale

× Utilisation uniquement pour le re-ranking

✓Sélection des informations locales influence l'interprétation

✓Recherche de corrélation locale pertinente (par objet)

État de l'Art

- **Relation de voisinage entre points d'intérêt**
  - Chen, Yap, Zhang : Discriminative Soft Bag-of-Visual Phrase for Mobile Landmark Recognition ; Transactions on Multimedia, 2014.
  - Torralba, Fergus, Weiss : Small Codes and Large Image Databases for Recognition ; Computer Vision and Pattern Recognition, 2008.

- **Mots visuels ⇒ « Phrase visuelle »**
- **Construction d'un voisinage de mots visuels (knn)**

× Lien entre les points clefs dans un voisinage de taille donnée sans considération d'objet réel

État de l'Art

- **Définition d'une région d'intérêt comme requête**
  - Chen : Efficient and Robust Image Ranking for Object Retrieval, University of Adelaide, 2013.
  - Chum, Mikulik, Perdoch, Matas : Total recall II: Query Expansion Revisited ; Computer Vision and Pattern Recognition, 2011.
  - Bursuc, Zaharia : Instance Search Task ; TRECVID Workshop, 2013.
  - Ali, Paar, Paletta : Semantic Indexing for Visual Recognition of Buildings ; International Symposium on Mobile Mapping Technology, 2007.

- **Spécification d'une ROI de l'objet d'intérêt**

× Uniquement en requête

× Définition manuelle

✓Résultats plus précis car l'information est plus pertinente

État de l'Art

- **Réduction et précision des mots clefs détectés**
  - Turcot, Lowe : Better matching with fewer features: The selection of useful features in large database recognition problems ; Conference on Computer Vision, 2009.
  - Weizman, Goldberger, Jacob: Urban-Area Segmentation Using Visual Words ; Geoscience and Remote Sensing Letters, 2009.
  - Li, Huang, Shao, Allinson : Building Recognition in Urban Environments: A Survey of State-of-the-Art and Future Challenges ; Information Sciences, 2014.
  - Li, Allinson : Building Recognition Using Local Oriented Features ; Transactions on Industrial Informatics, 2013.
  - Ionescu, Benois-Pineau, Piatrik, Quenot : Fusion in Computer Vision: Understanding Complex Visual Content ; Computer Vision and Pattern Recognition, 2014.

- **Vocabulaire de patchs, regroupement des mots par grappe, filtrage des mots redondants**

× Risque de répétition du même mot sur un bâtiment

× Risque de réduire le caractère discriminant de certains mots

✓ Sélection uniquement de l'information utile

État de l'Art

- **Réseaux neuronaux par apprentissage profond**
  - Simonyan, Vedaldi, Zisserman : Deep Fisher Networks for Large-Scale Image Classification ; Advances in Neural Information Processing Systems, 2013.
  - Bezak : Building Recognition System Based on Deep Learning ; Artificial Intelligence and Pattern Recognition, 2016.
  - Wang, Zhang, Li, Zhang, Lin : Cost-Effective Active Learning for Deep Image Classification ; Circuits and Systems for Video Technology, 2016.

- **Parallélisation du travail de représentation d'une image par portion**

× Problème de mémoire

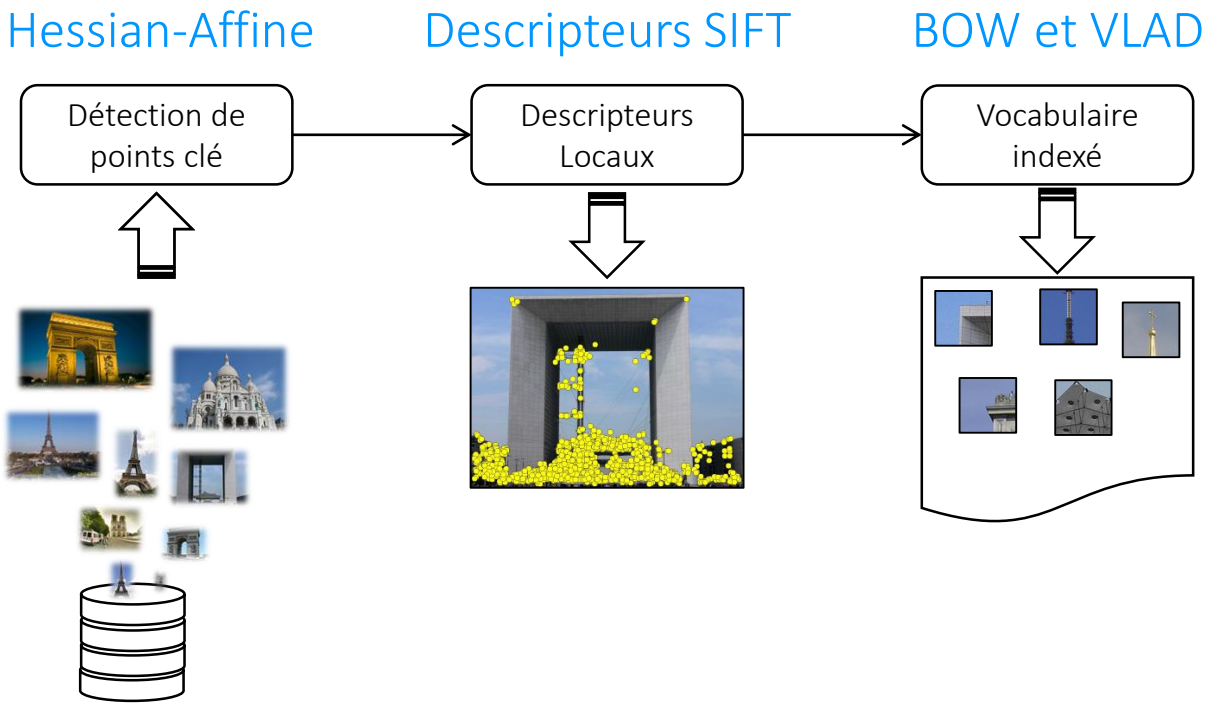
× Étape d'entraînement critique : nécessite un grand nombre de données

État de l'Art

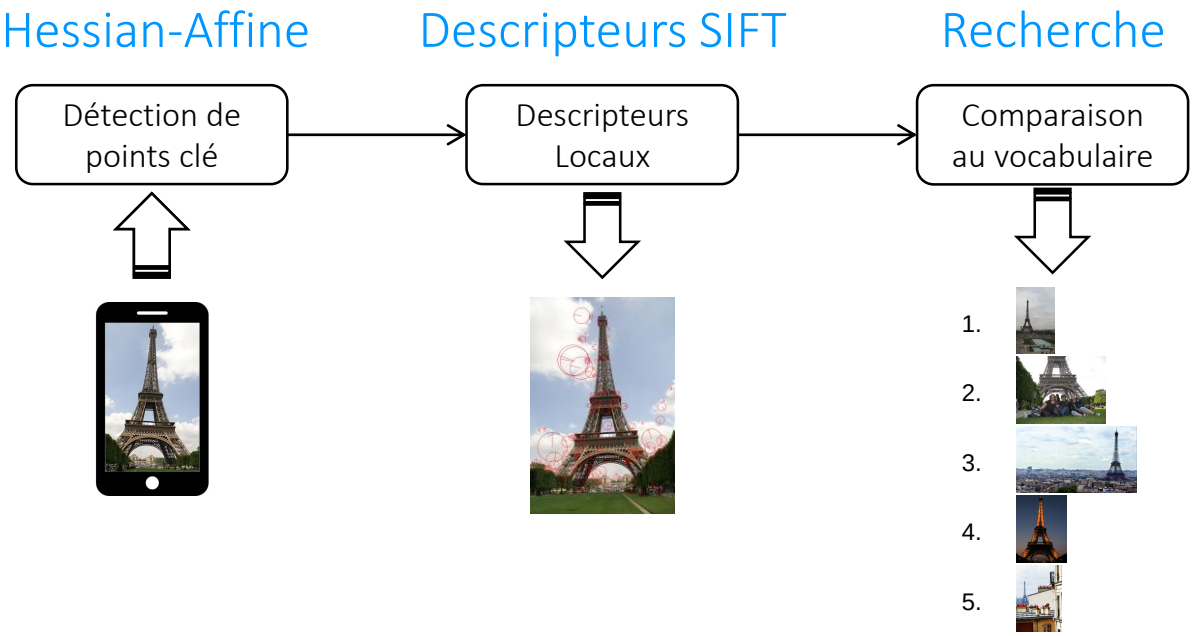
- L'information extraite d'une image reste LOCALE et INDÉPENDANTE
- Pas de CONTEXTE SÉMANTIQUE
- Dépend d'un ENTRAÎNEMENT spécifique et IMPORTANT
- Problèmes de GÉNÉRALISATION
- VERSATILITÉ des images représentant un bâtiment/monument dans une scène EXTÉRIEURE et PUBLIQUE
- Pas de distinction entre les différents mots clefs extraits

Requête d’image

Construction d'un vocabulaire (OFFLINE)

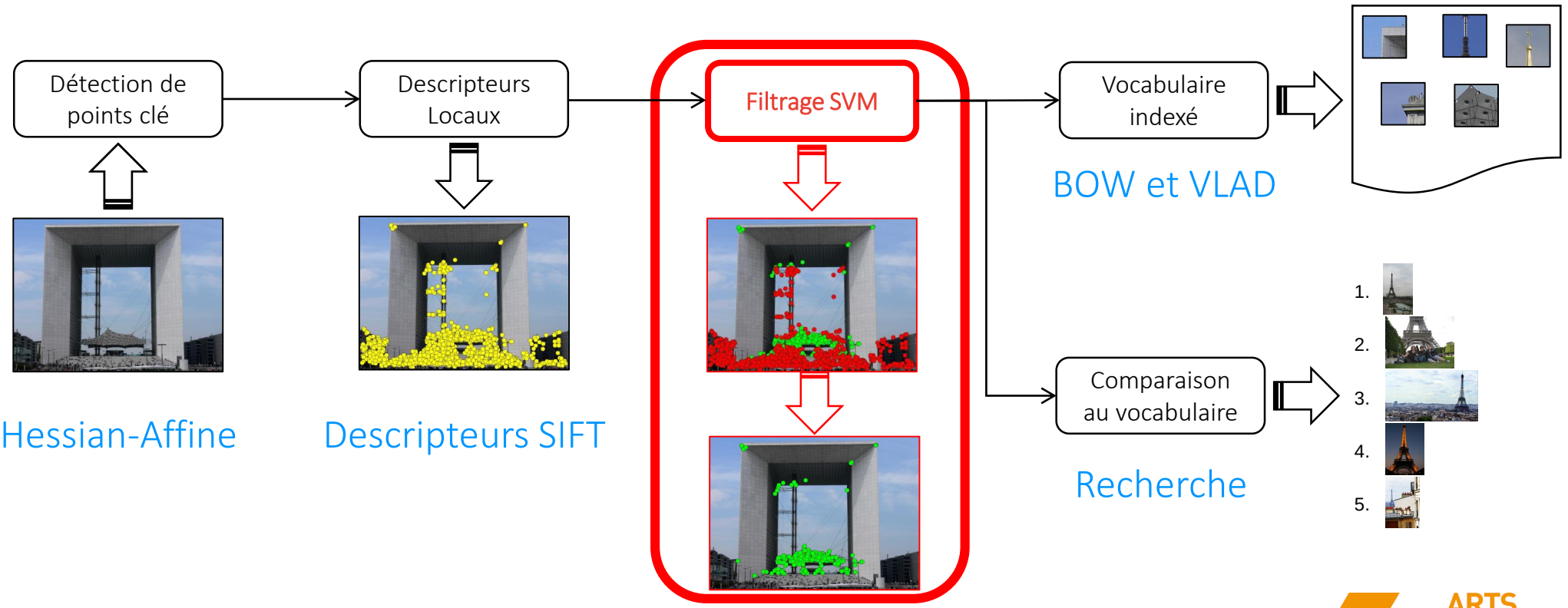


Résultat d'une image requête (ONLINE)



Requête d'image

- Approche de description par points d'intérêt avec FILTRAGE BINAIRE des POINTS D'INTÉRÊTS LOCAUX en **bâtiment**/**non-bâtiment**



Requête d'image

- Une approche de CLASSIFICATION GLOBALE des descripteurs locaux
- Une approche de CLASSIFICATION PAR MODÈLES SVM MULTIPLES ADAPTÉS à chaque catégorie de bâtiment
- Une technique de VÉRIFICATION GÉOMÉTRIQUE pour la prise en compte du contexte spatial
- Validation expérimentale sur deux bases de données publiquement disponibles (Paris 6k et Oxford 5k)

Plan

- Introduction
- Etat de l’art et contributions
- Bases de données d’expérimentation
- Classification globale des descripteurs locaux
- Modèles SVM adaptés par classe de bâtiments
- Vérification et correction géométrique
- Recherche par similarité : résultats expérimentaux
- Conclusion

Bases de donnée d’expérimentation

- Paris 6k
  - 6 412 images issues de 11 monuments touristiques différents à Paris :
    1. L’arche de la Défense,
    2. La Tour Eiffel,
    3. L'Hôtel des Invalides,
    4. Le Louvre,
    5. Le Moulin Rouge,
    6. Le Musée d'Orsay,
    7. Notre Dame,
    8. Le Panthéon,
    9. Le musée Pompidou,
    10. Le Sacré Cœur et
    11. L'Arc de Triomphe.



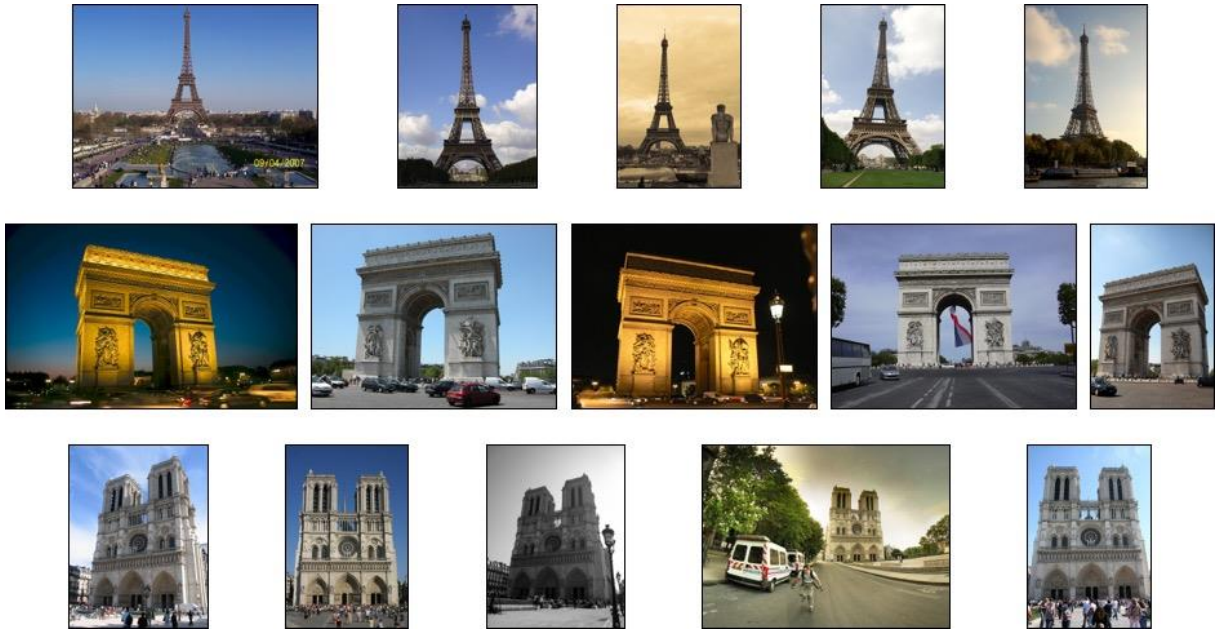







Bases de donnée d’expérimentation

- Paris 6k : Vérité terrain
  - 5 images de chaque catégories définies en requête
  - Une partie est définie comme images correctement détectées
  - L’autre comme images erronées



Bases de donnée d’expérimentation

- Oxford 5k
  - 5 063 images représentant 11 sites d'intérêt touristique à Oxford :
    1. All Souls,
    2. Ashmolean,
    3. Balliol,
    4. Bodleian,
    5. Christ Church,
    6. Cornmarket,
    7. Hertford,
    8. Keble,
    9. Magdalen,
    10. Pitt Rivers et
    11. Radcliffe Camera











Plan

- Introduction
- Etat de l'art et contributions
- Bases de données d'expérimentation
- Classification globale des descripteurs locaux
- Modèles SVM adaptés par classe de bâtiments
- Vérification et correction géométrique
- Recherche par similarité : résultats expérimentaux
- Conclusion

Descripteur local d'image

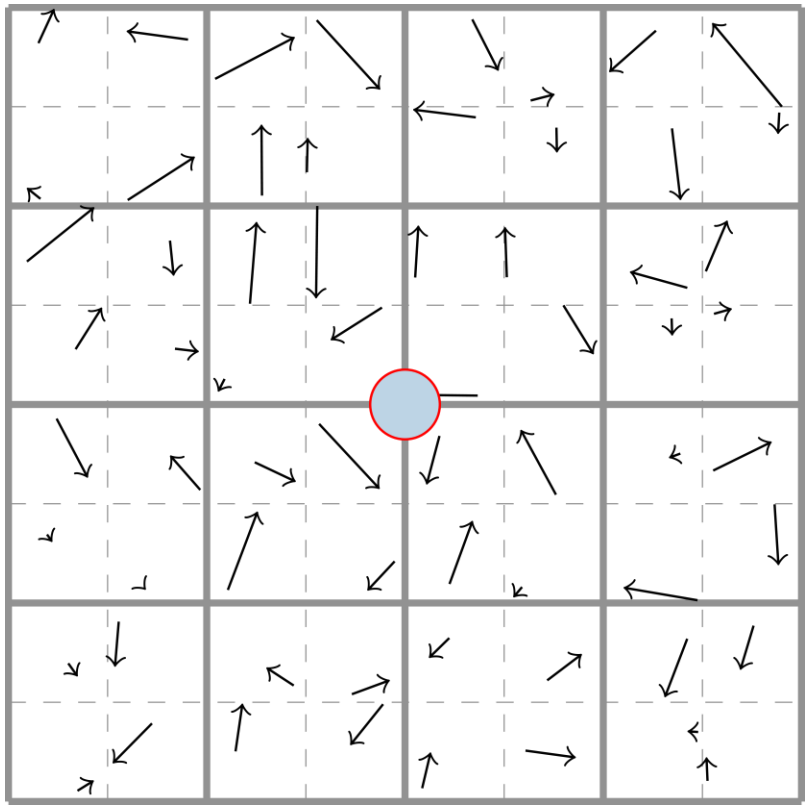
- SIFT : Scale Invariant Feature Transform
- ∀ point d'intérêt détecté, GRADIENT D'IMAGE $\mathcal{L}$ (≈ direction de la variation d'intensité)
- $\mathcal{L}$ est défini par son AMPLITUDE m et son ORIENTATION θ

$$\mathcal{L}(x,y): \begin{cases} m(x,y) = \sqrt{\Delta_x^2 + \Delta_y^2} \\ \theta(x,y) = \tan^{-1}\left(\frac{\Delta_y}{\Delta_x}\right) \end{cases}$$

$$\text{avec} \begin{cases} \Delta_x = \mathcal{L}(x+1,y) - \mathcal{L}(x-1,y) \\ \Delta_y = \mathcal{L}(x,y+1) - \mathcal{L}(x,y-1) \end{cases}$$

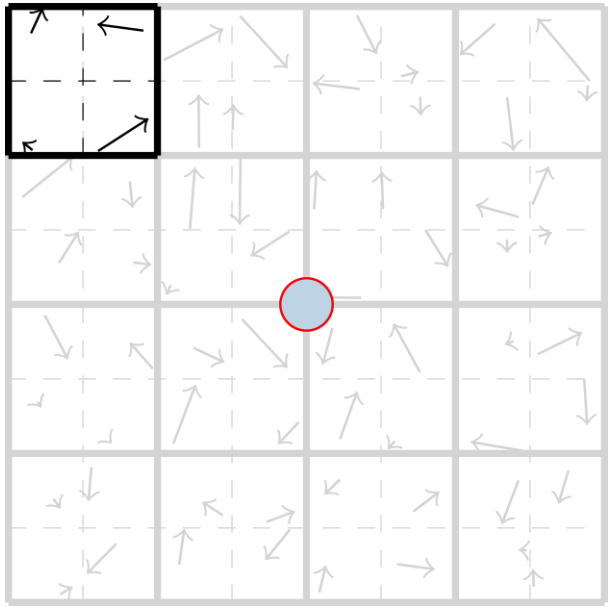
Descripteur local d'image

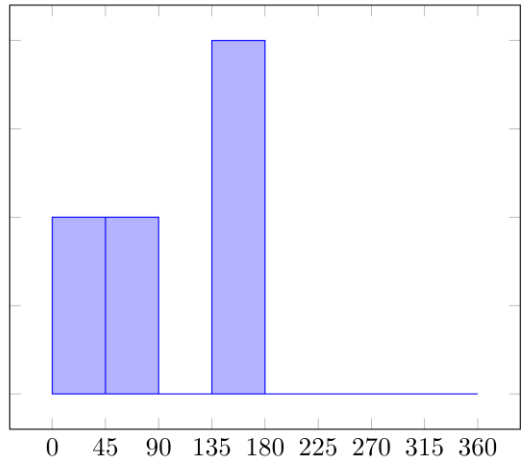
- Voisinage de (16 × 16)
- 16 sous-zones de 4 × 4



Descripteur local d'image

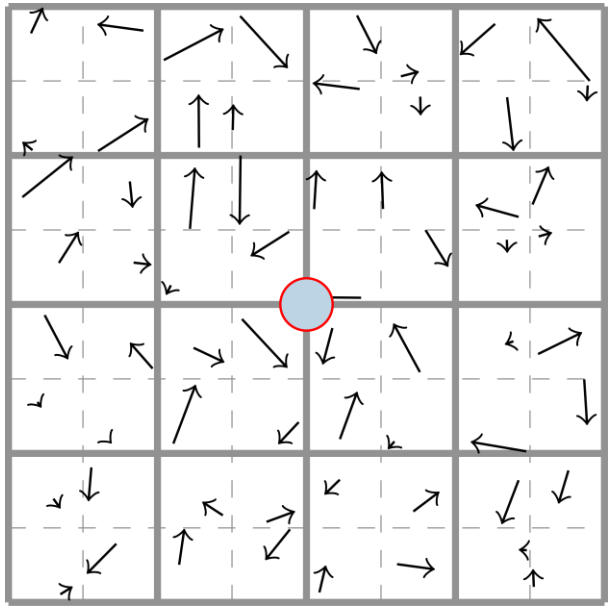
- Voisinage de (16 × 16)
- 16 sous-zones de 4 × 4
- HISTOGRAMME (8 intervalles) de GRADIENT ORIENTÉ (HOG) pour chaque sous-zone

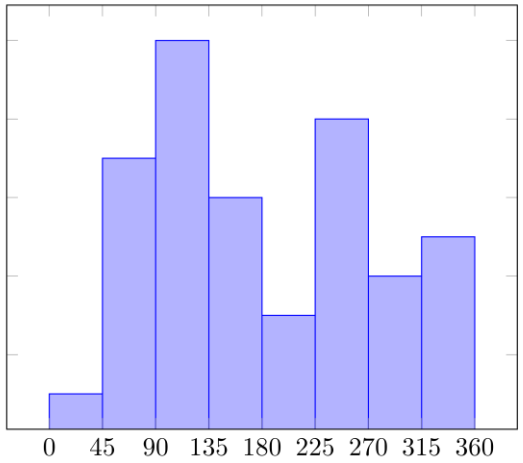




Descripteur local d'image

- Voisinage de (16 × 16)
- 16 sous-zones de 4 × 4
- HISTOGRAMME (8 intervalles) de GRADIENT ORIENTÉ (HOG) pour chaque sous-zone
- HOG pour chacune des 16 sous-zones ⇒ 16 × 8 = 128
- Un mot visuel ⇔ vecteur de 128 COMPOSANTES





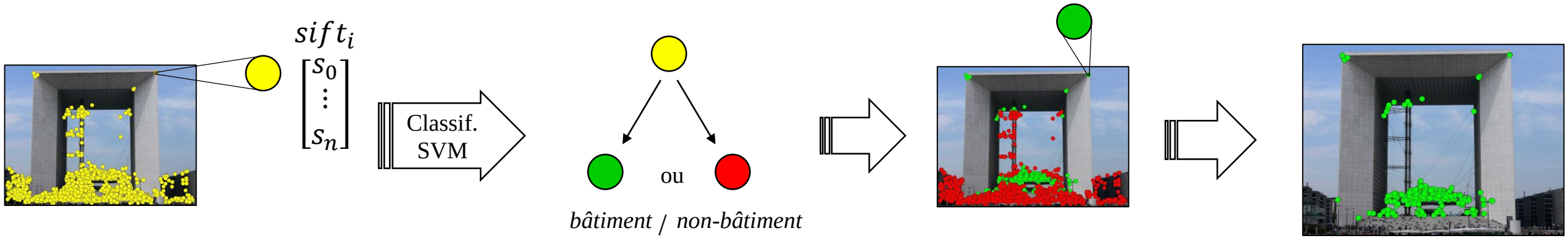
Détecteur de points d'intérêt

- Point d'intérêt
  - Maximum de variation de GRADIENT D'IMAGE (≈ direction de la variation d'intensité)
  - Maximum d'information DISCRIMINANTE
- Détecteur Hessian-affine : Matrice Hessienne normalisée
  - Analyse dans un ESPACE D'ÉCHELLE GAUSSIEN
  - Bonnes propriétés de RÉPÉTABILITÉ et invariance para rapports à l'échelle et à la pose

$$\mathcal{H}_\sigma(x,y) = \begin{bmatrix} \frac{\partial I_\sigma(x,y)}{\partial x^2} & \frac{\partial I_\sigma(x,y)}{\partial x \partial y} \\ \frac{\partial I_\sigma(x,y)}{\partial y \partial x} & \frac{\partial I_\sigma(x,y)}{\partial y^2} \end{bmatrix}$$

Classification globale des descripteurs locaux

- Principe proposé : filtrage SVM (Support Vector Machines) des points d'intérêt
  - Classification supervisée dans l'espace des descripteurs SIFT



Classification Support Vector Machine

- Entraînement : optimisation $\min_a \|a\|^2$ tel que $y_i(ax_i + b) \geq 1$
- $e_i \geq 0$: distance du point i à la FRONTIÈRE
- Frontière « souple » :

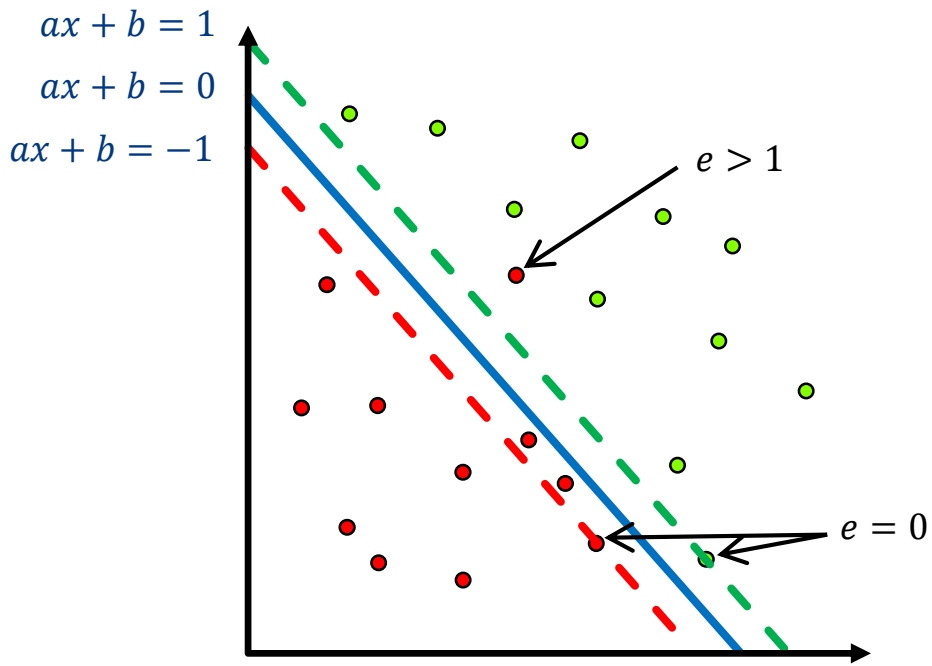
$$\min_a \|a\|^2 + C \sum_{i=1}^{N} e_i$$

- Tel que $y_i(ax_i + b) \geq 1 - e_i$

- Equations des NOYAUX SVM :
  - Linéaire : $k(u, v) = u' \times v$
  - Polynomial : $k(u, v) = (\gamma\, u' \times v + r)^{\delta}$
  - Radial : $k(u, v) = \exp(-\gamma\, |u - v|^2)$

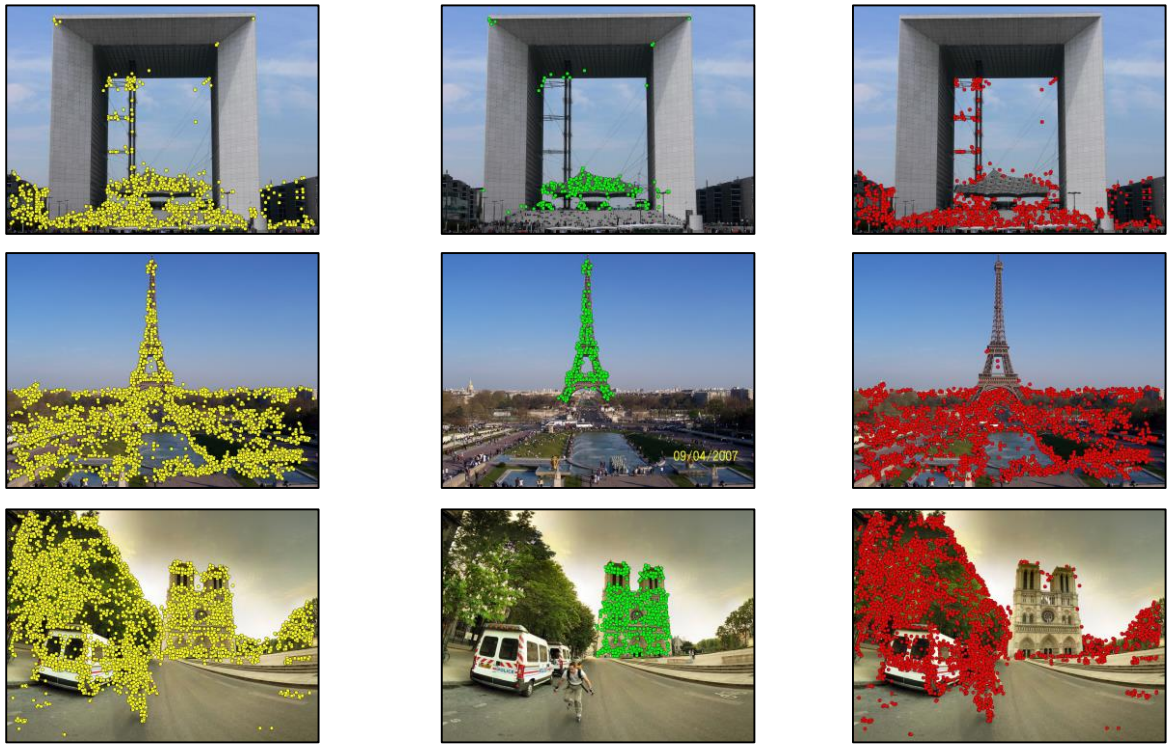
avec C le paramètre de PÉNALISATION

avec γ le paramètre d'INFLUENCE



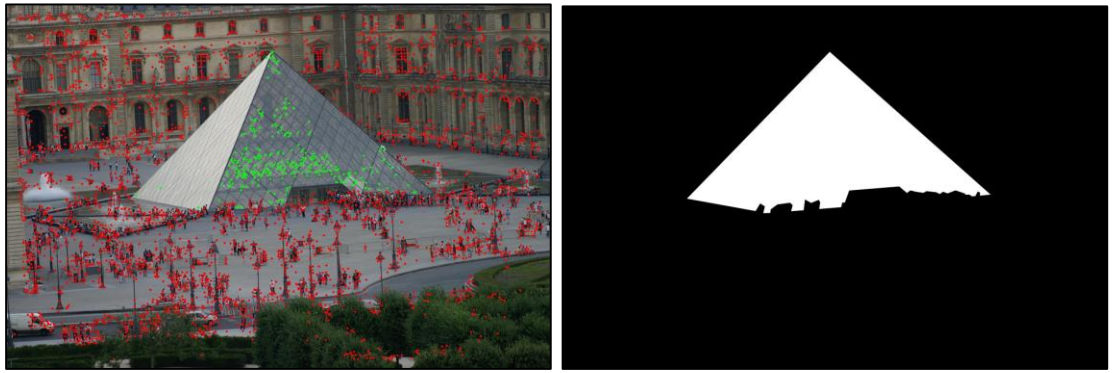
Classification globale des descripteurs locaux

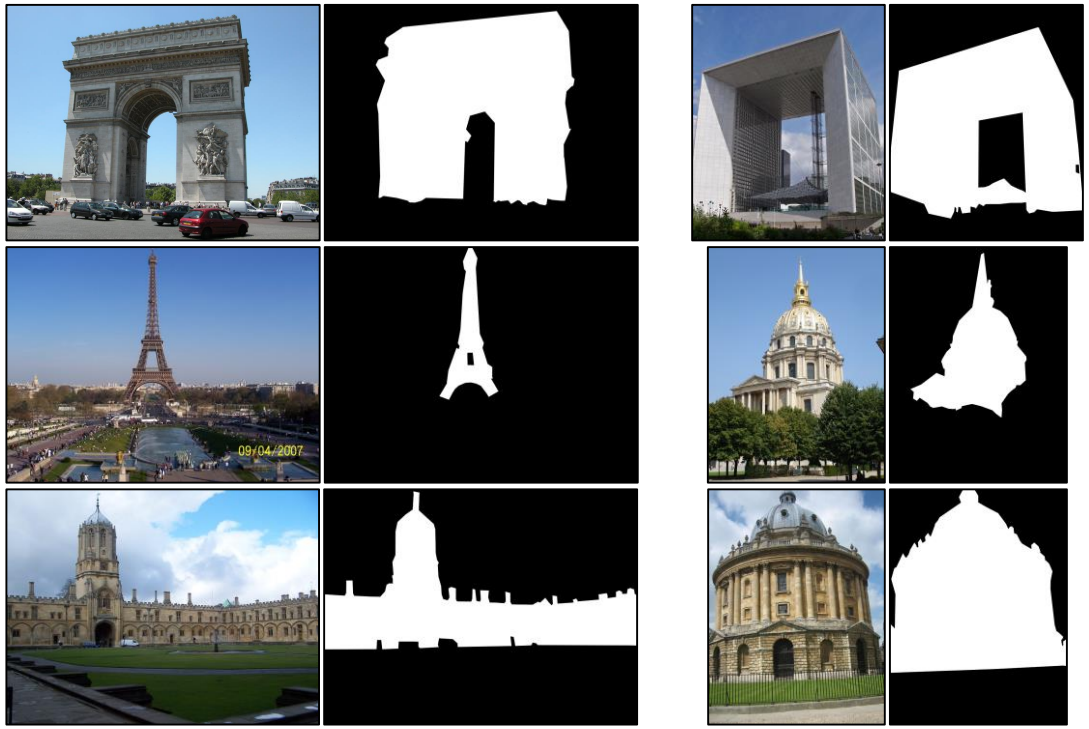
- Difficultés
  - PLUSIEURS MILLIERS de points clefs peuvent être détectés pour une image
  - Déterminer les bons PARAMÈTRES de classification et éviter l'OVERFITTING



Classification globale des descripteurs locaux

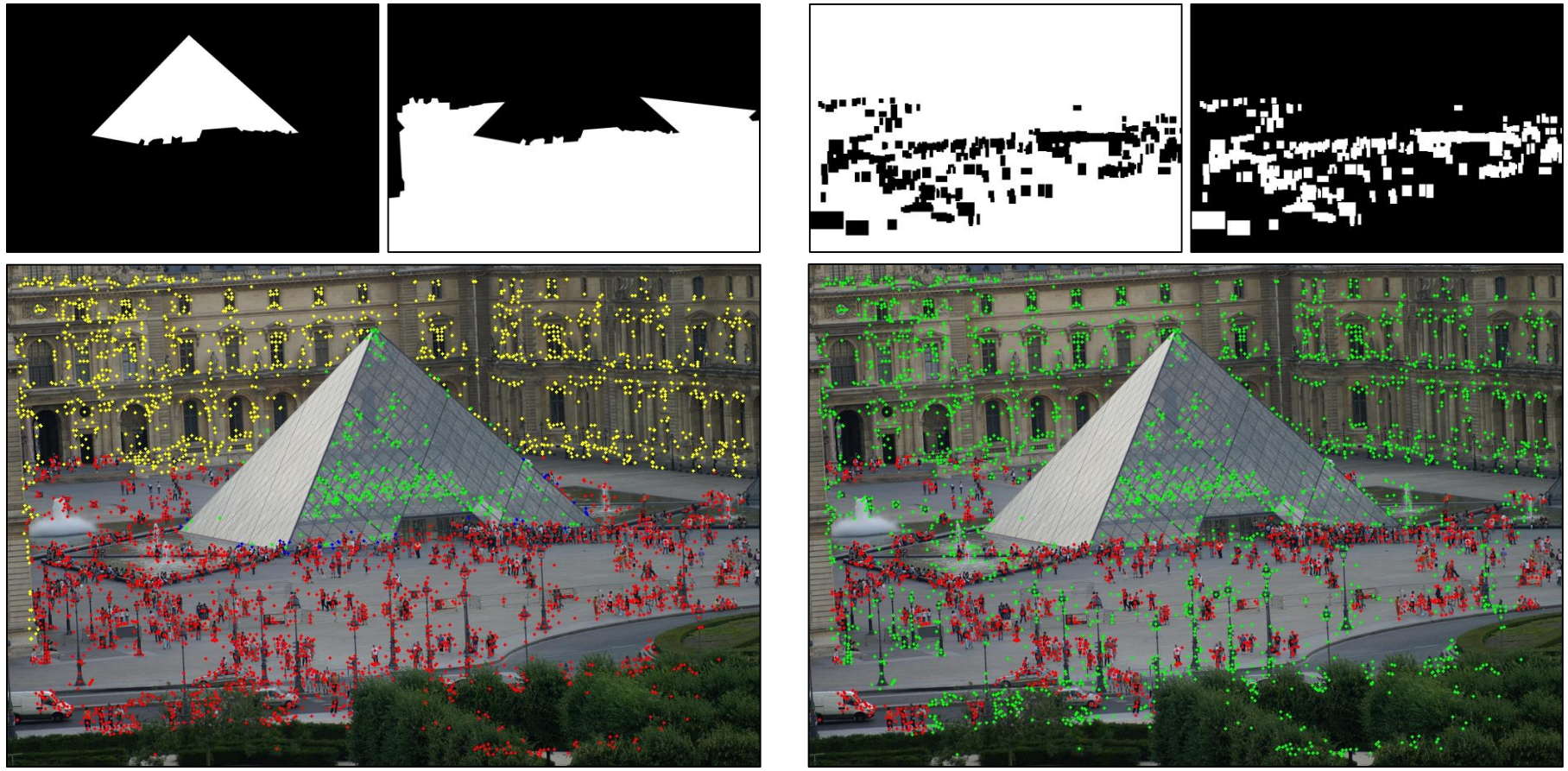
- Protocole d'entraînement
  - 5 images requêtes par catégorie de bâtiment : 55 images
  - Masques binaires de sélection de données





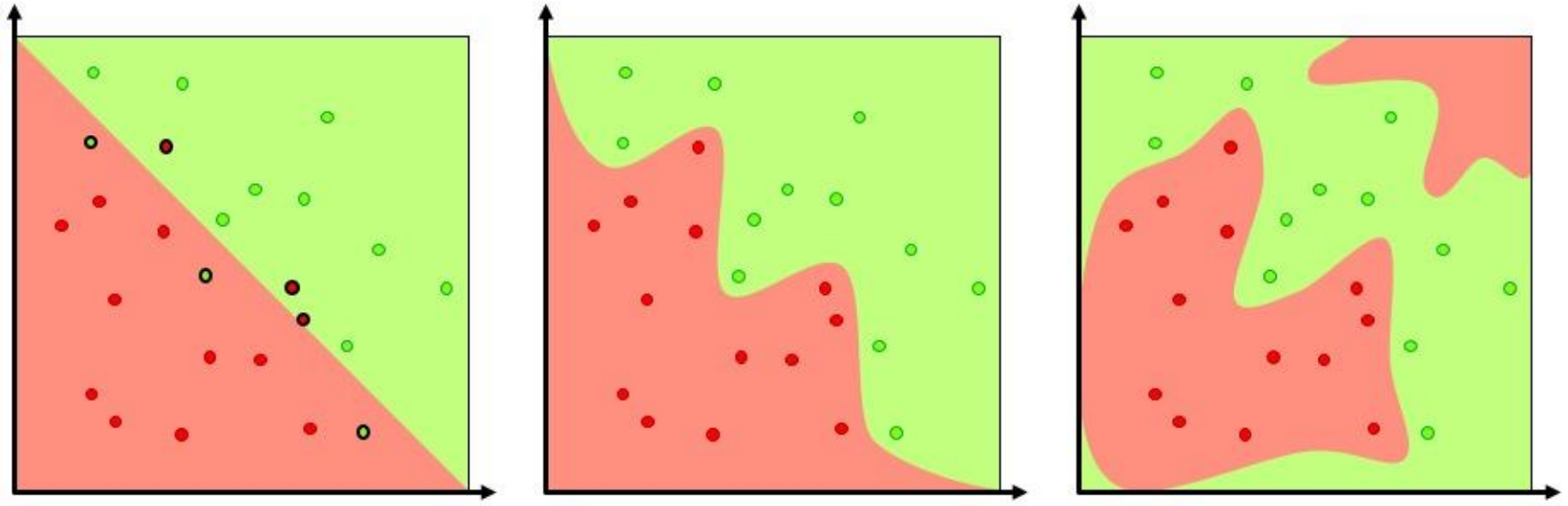
Classification globale des descripteurs locaux

- Différentes approches de sélection par masque testées



Classification globale des descripteurs locaux

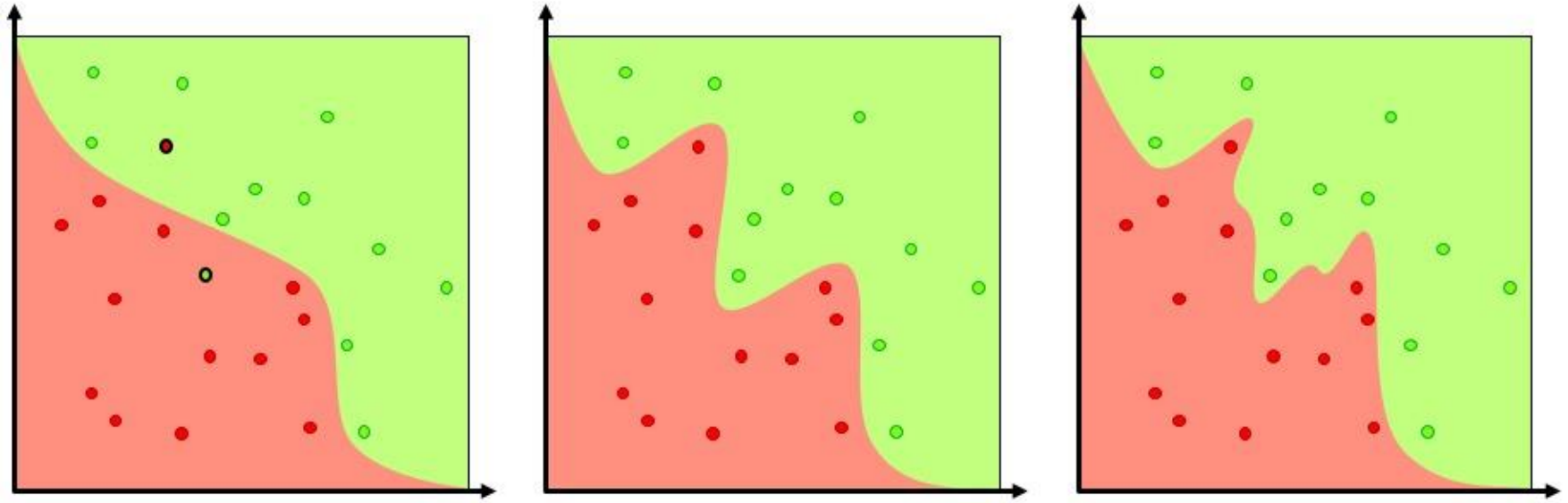
- Paramètres d'apprentissage : LE NOYAU



Linéaire Polynomial Radial

Classification globale des descripteurs locaux

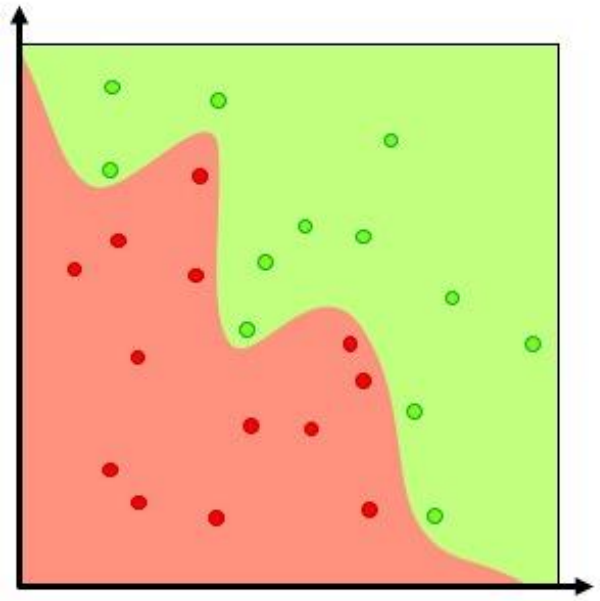
- Paramètres d'apprentissage : LE DEGRÉ δ DU POLYNÔME



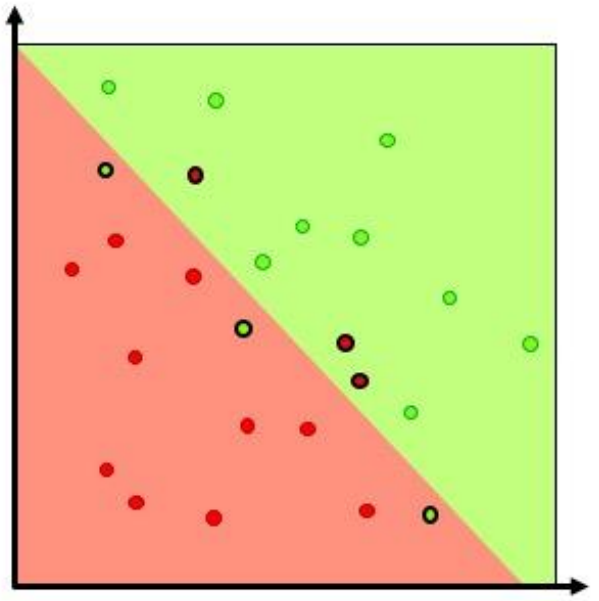
$\delta = 3$ $\delta = 5$ $\delta = 7$

Classification globale des descripteurs locaux

- Paramètres d'apprentissage : PARAMÈTRE DE PÉNALISATION C



Valeur élevée



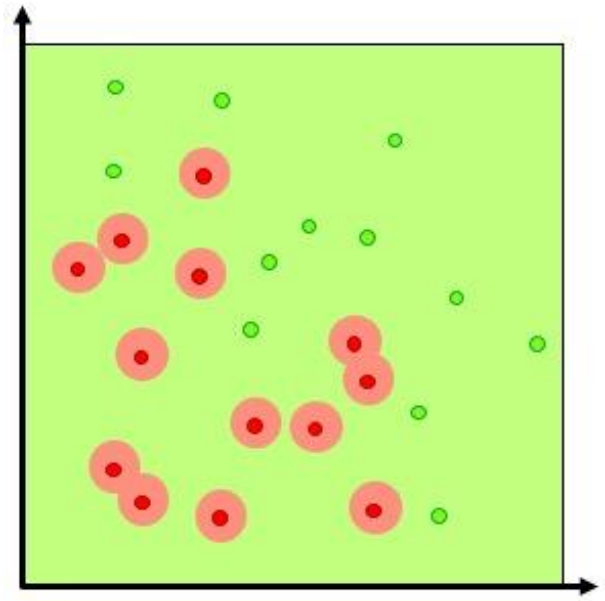
Valeur faible

$$\min_a \|a\|^2 + C \sum_{i=1}^{N} e_i$$

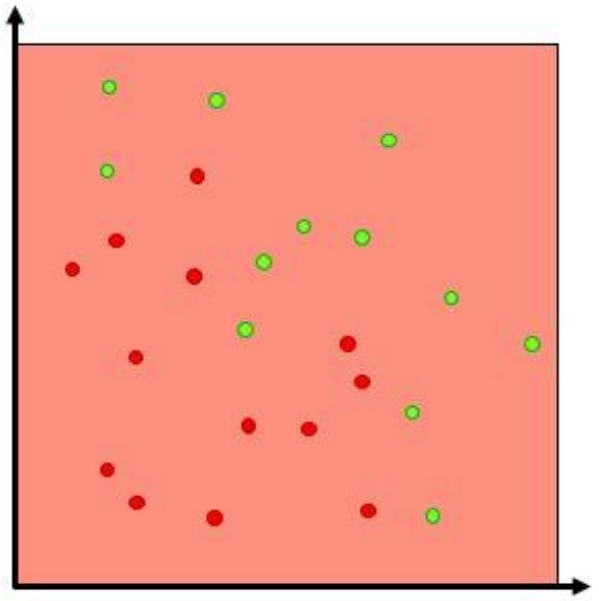
Classification globale des descripteurs locaux

- Paramètres d'apprentissage : PARAMÈTRE D'INFLUENCE γ

$$k(u, v) = (\gamma\, u' \times v + r)^{\delta}$$



Valeur faible



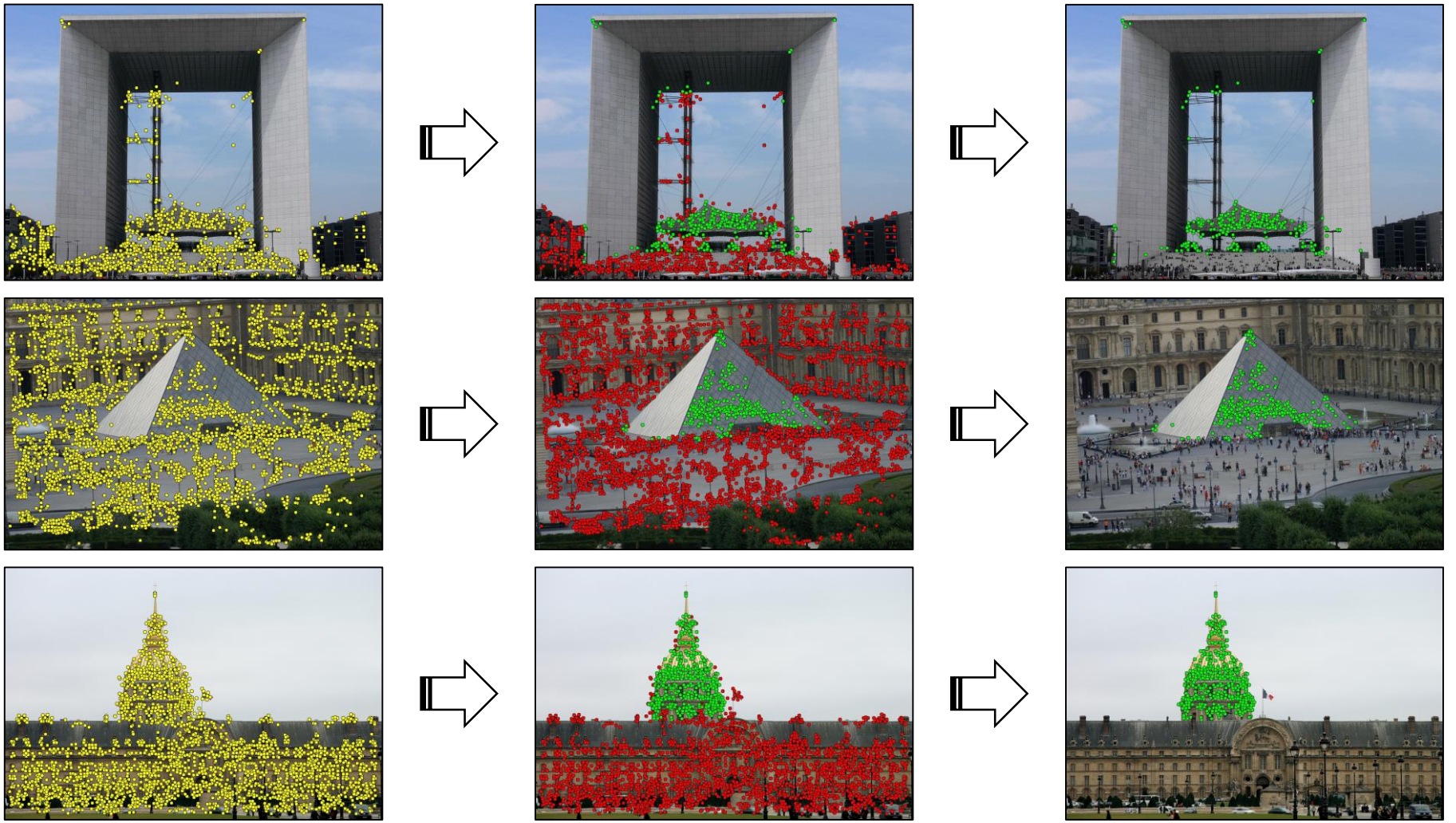
Valeur élevée

Classification globale des descripteurs locaux

- Paramètres retenus pour l'entraînement des modèles SVM :
  - Noyau POLYNOMIAL
  - Degré du polynôme $\delta = 5$
  - Paramètre d'influence $\gamma = 1$
  - Paramètre de pénalisation $C = 10$

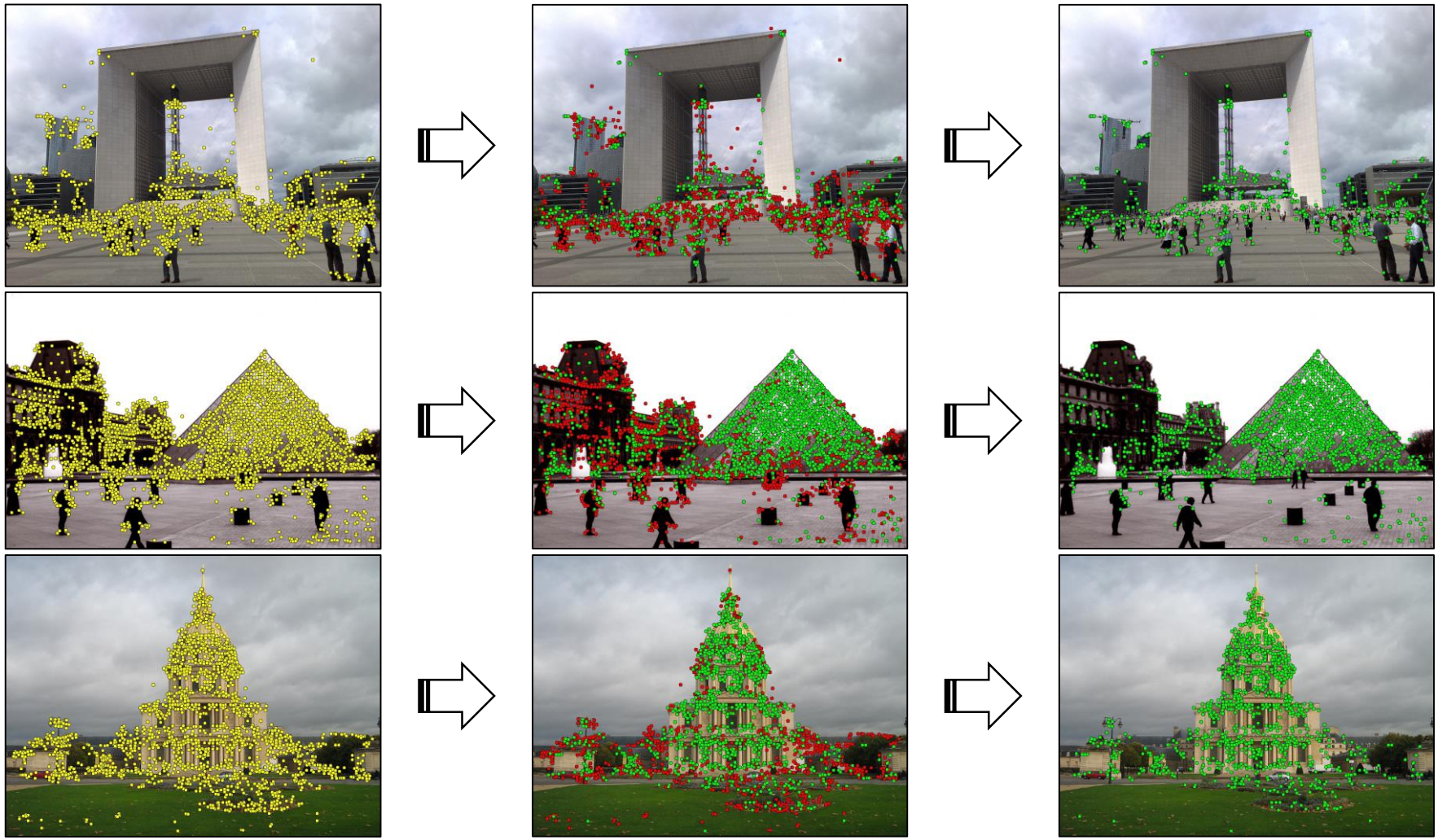
Classification globale des descripteurs locaux

Prédiction sur des images d'entraînement



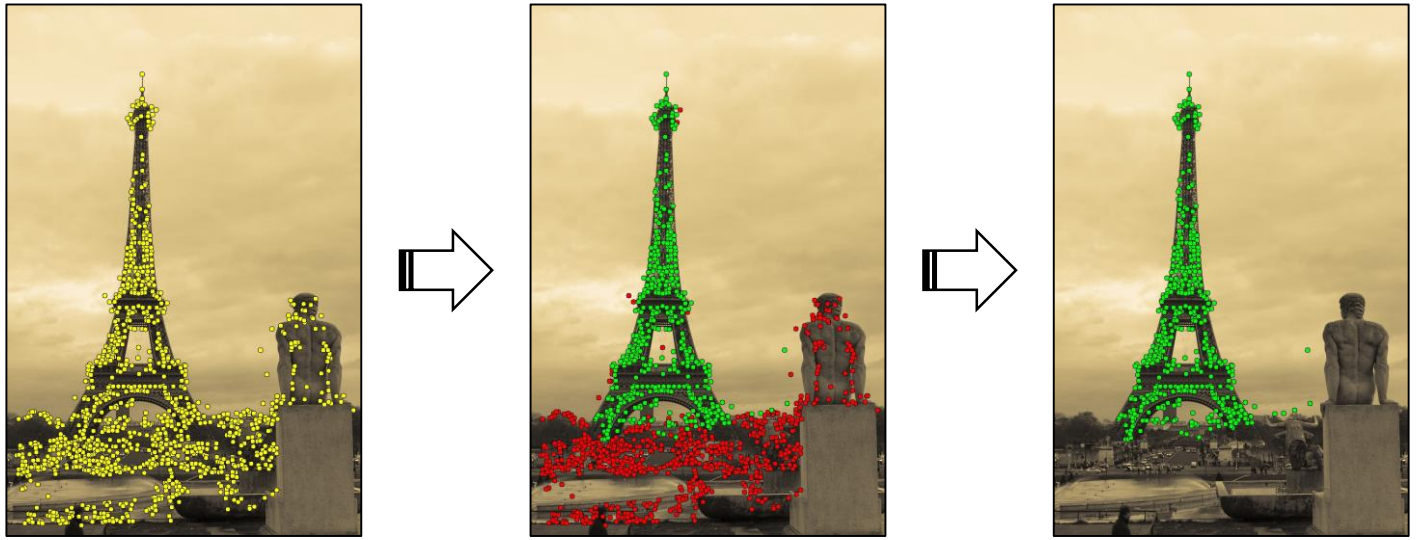
Classification globale des descripteurs locaux

Prédiction sur des images de test (GÉNÉRALISATION)

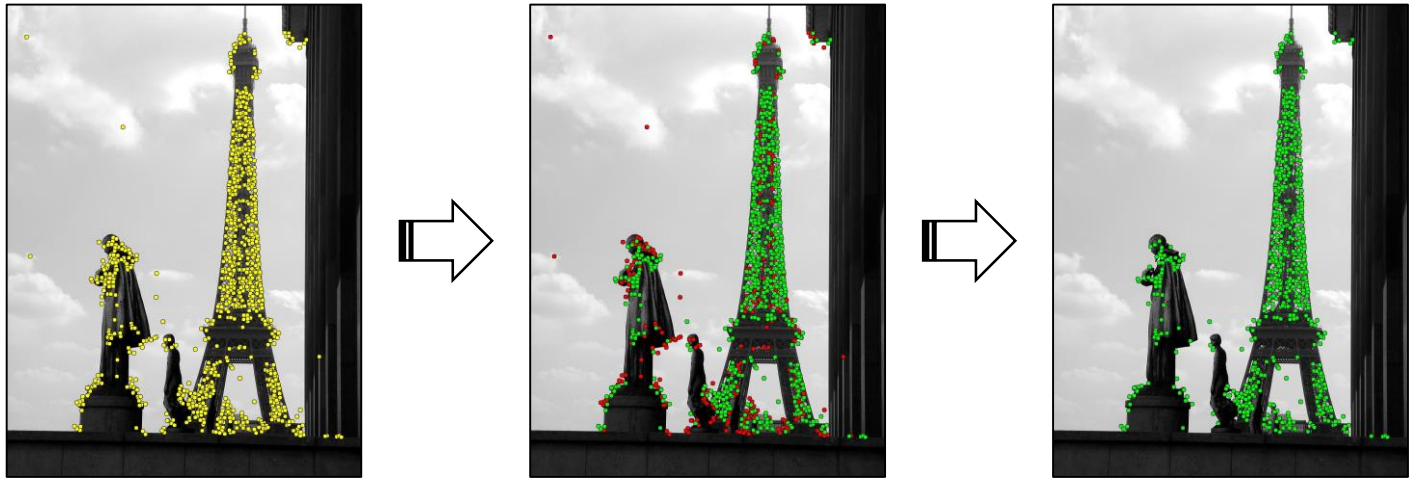


Classification globale des descripteurs locaux

Prédiction sur une image d'entraînement

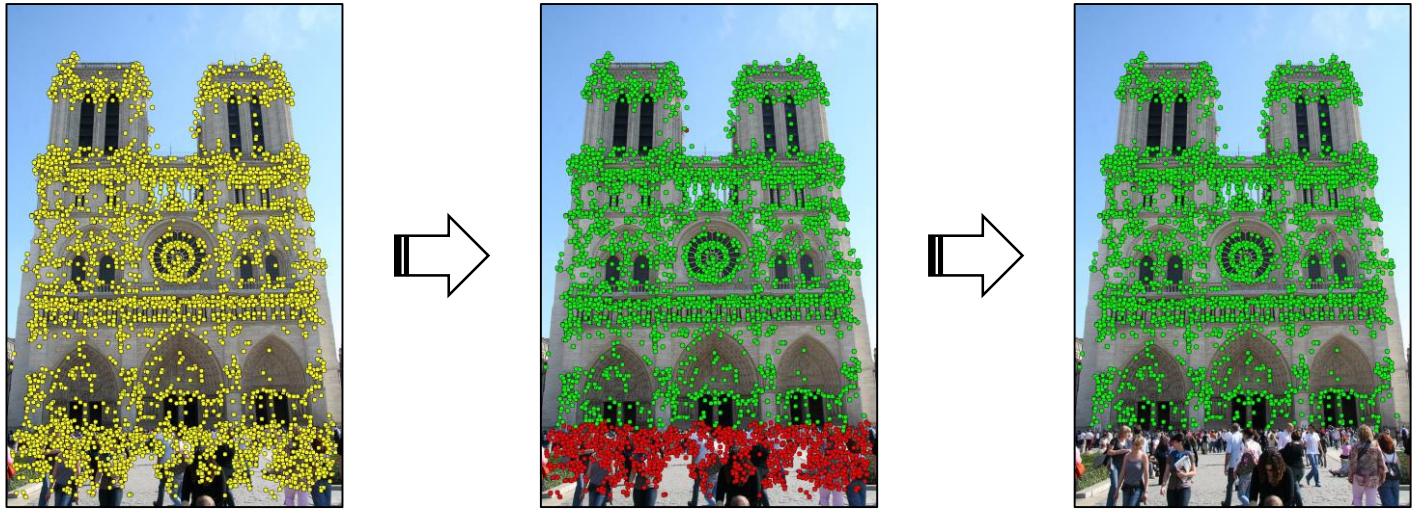


Prédiction sur une image de test (GÉNÉRALISATION)

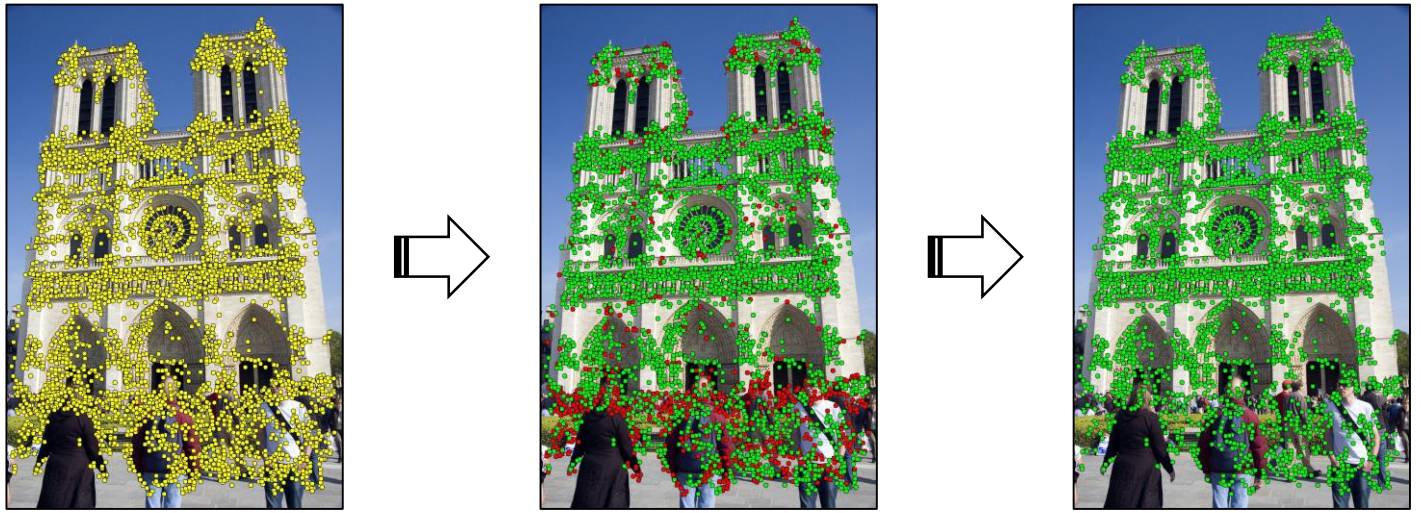


Classification globale des descripteurs locaux

Prédiction sur une image d'entraînement

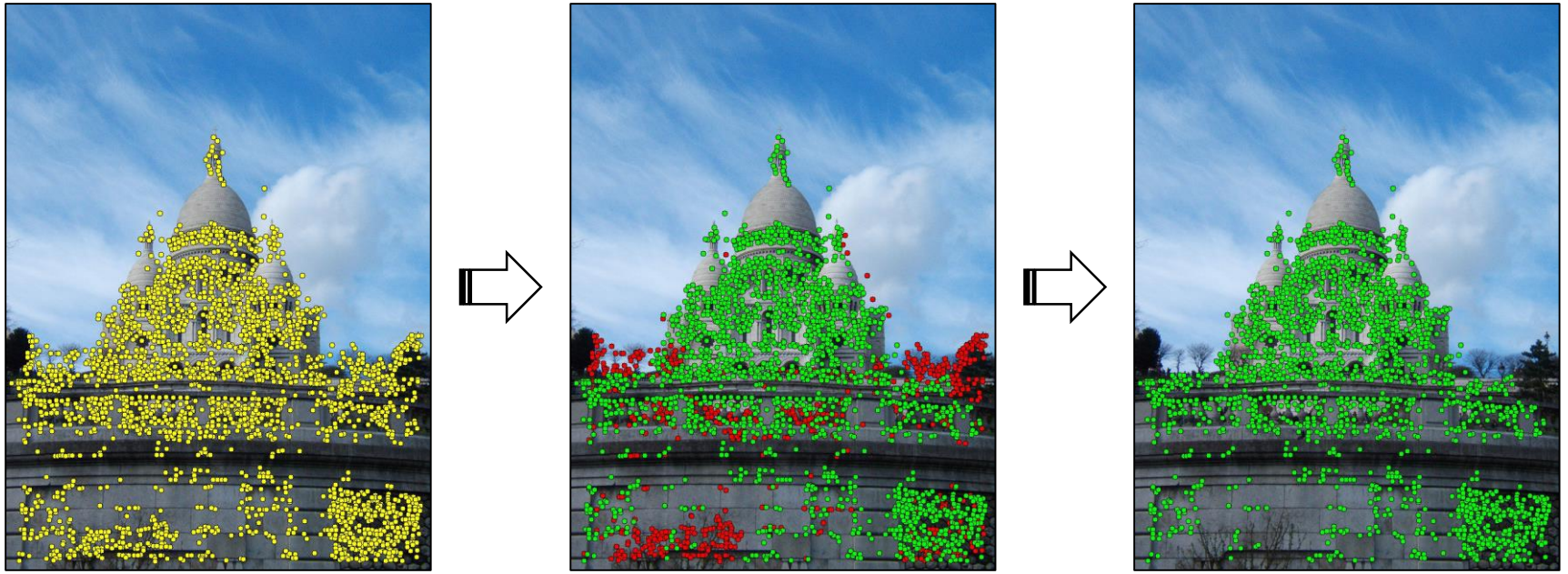


Prédiction sur une image de test (GÉNÉRALISATION)



Classification globale des descripteurs locaux

Prédiction sur une image d'entraînement



Prédiction sur une image de test (GÉNÉRALISATION)

Classification globale des descripteurs locaux

- Analyse : bilan et limitation
  - La méthode de classification binaire donne des résultats prometteurs
  - Distinctions des classes SÉMANTIQUES *bâtiment* et *non-bâtiment*
  - Problématique de GÉNÉRALISATION des prédictions aux images tests

Catégories	Faux positifs par image	Descripteurs *bâtiment* par image	Ratio
La Défense	373	612	0,609
Tour Eiffel	425	1041	0,408
Invalides	458	1027	0,446
Louvre	560	1227	0,457
Moulin Rouge	518	1066	0,486
Musée d'Orsay	1198	1769	0,677
Notre Dame	250	2013	0,124
Panthéon	200	1115	0,180
Pompidou	383	3025	0,127
Sacré Cœur	285	1243	0,230
Arc de Triomphe	464	1295	0,359
Moyenne globale	**465**	**1403**	**0,373**

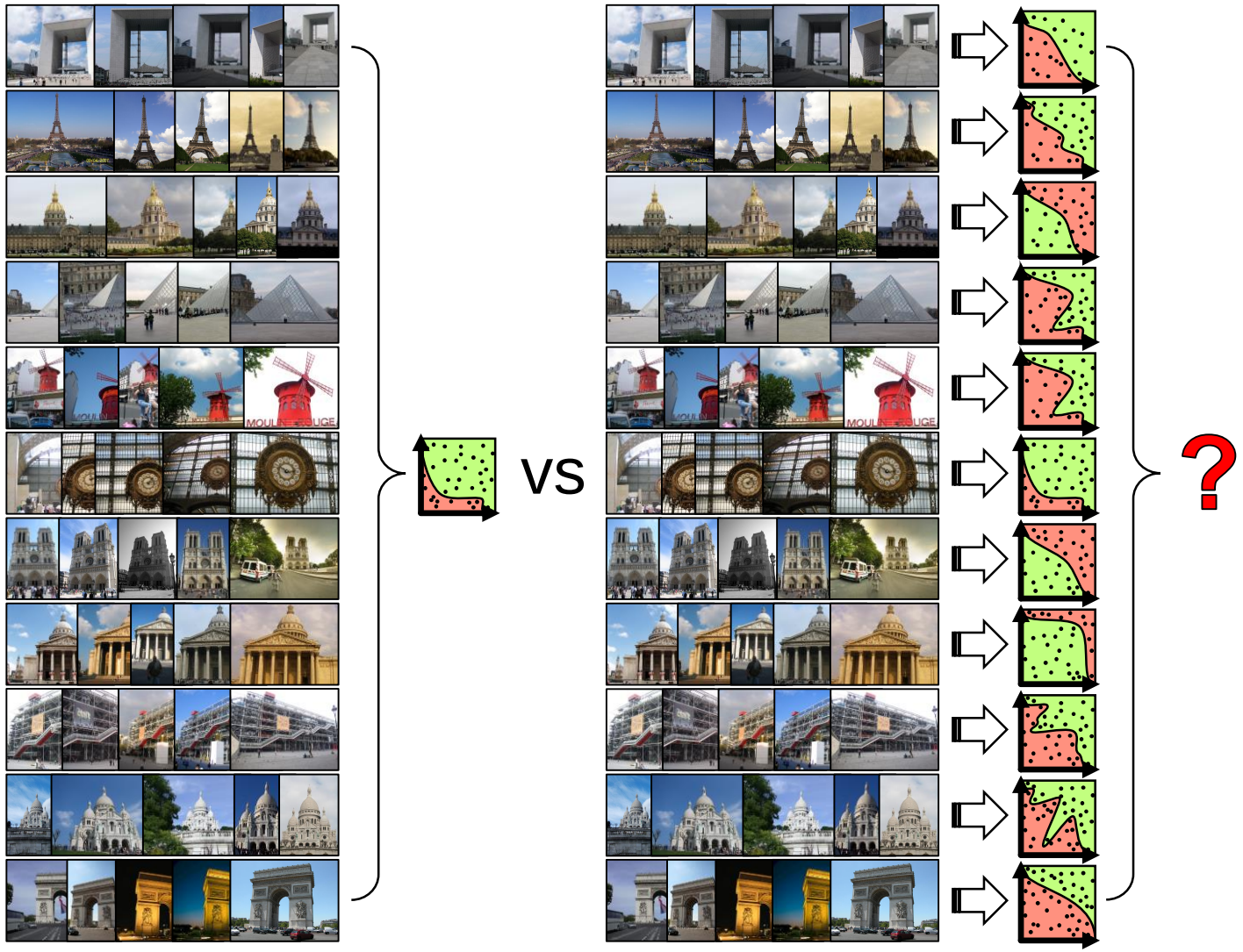
Plan

- Introduction
- Etat de l'art et contributions
- Bases de données d'expérimentation
- Classification globale des descripteurs locaux
- Modèles SVM adaptés par classe de bâtiments
- Vérification et correction géométrique
- Recherche par similarité : résultats expérimentaux
- Conclusion

Modèles SVM adaptés par classe de bâtiments

- Un entraînement spécifique SVM par CATÉGORIE DE BÂTIMENT
  - Résulte en DIFFÉRENTS modèles SVM applicables

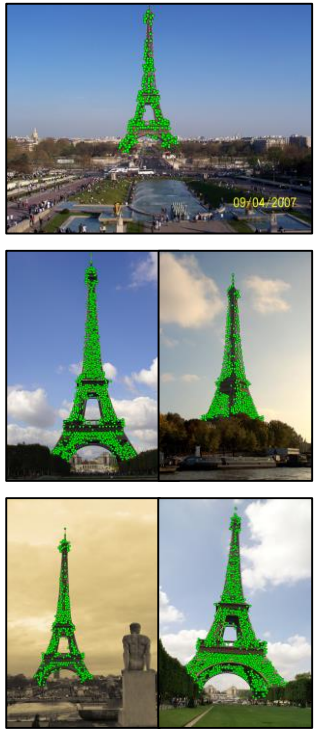
- Problématique de FUSION sur image inconnue *a priori*



Modèles SVM adaptés par classe de bâtiments

- Protocole d'entraînement
- Données indépendantes par catégorie de bâtiment
  - Discrimination d'un bâtiment par rapport aux autres
  - Classe *bâtiment* restrictive

Données d'entraînement *bâtiment*



Données d'entraînement *non-bâtiment*

Modèles SVM adaptés par classe de bâtiments

- Fusion
  - Pour chaque point d'intérêt : SCORE DE CONFIANCE de la prédiction SVM dépendant du modèle SVM associé

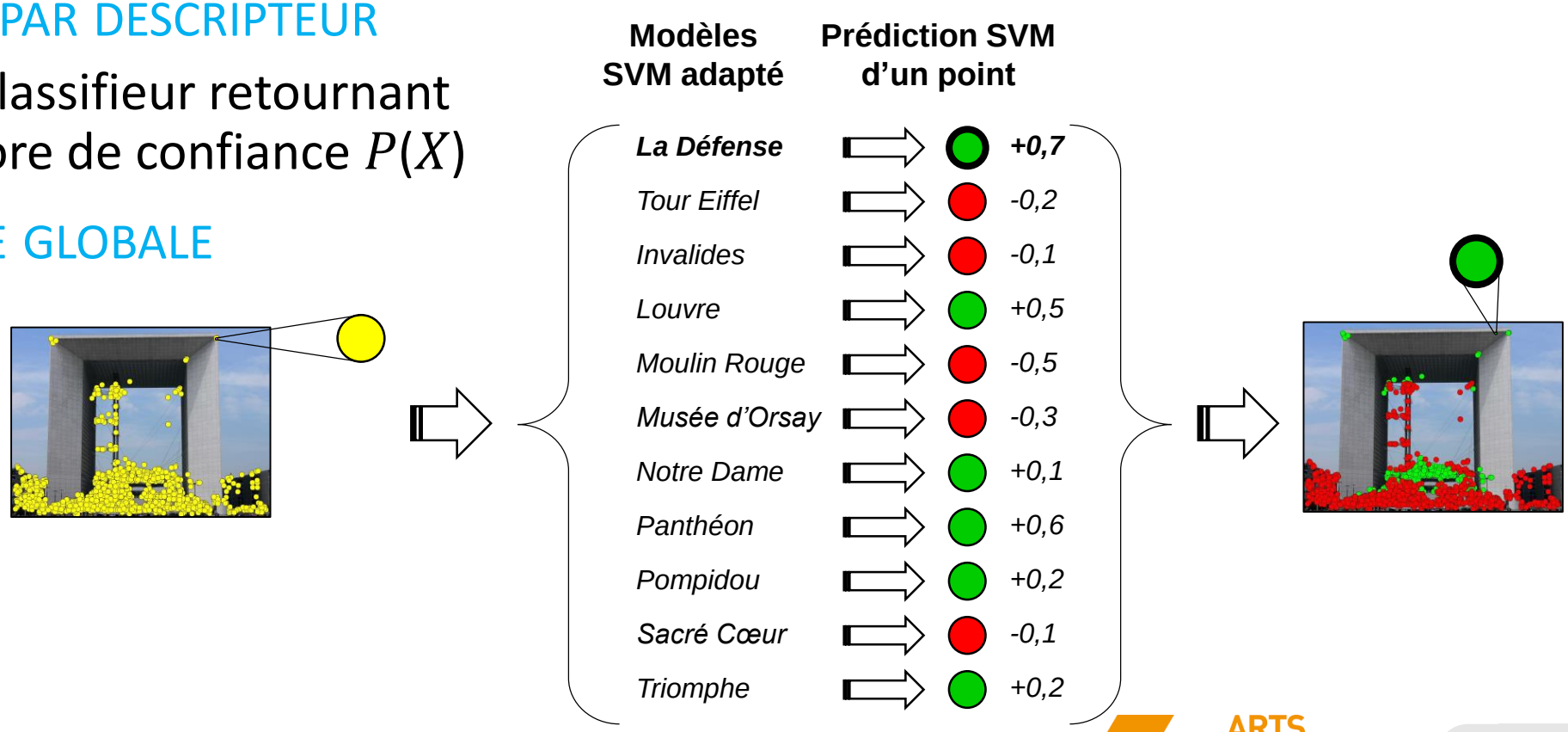
$$P(X) = \frac{1}{1 + \exp(-|\Delta_X|)}$$

Δ_X : Distance du point de descripteur X à la marge

- 1 modèle SVM entraîné / catégorie de bâtiment
- PROBLÉMATIQUE : Comment choisir le modèle SVM adapté ?
  A quel échelle ? Descripteur ou Image ?

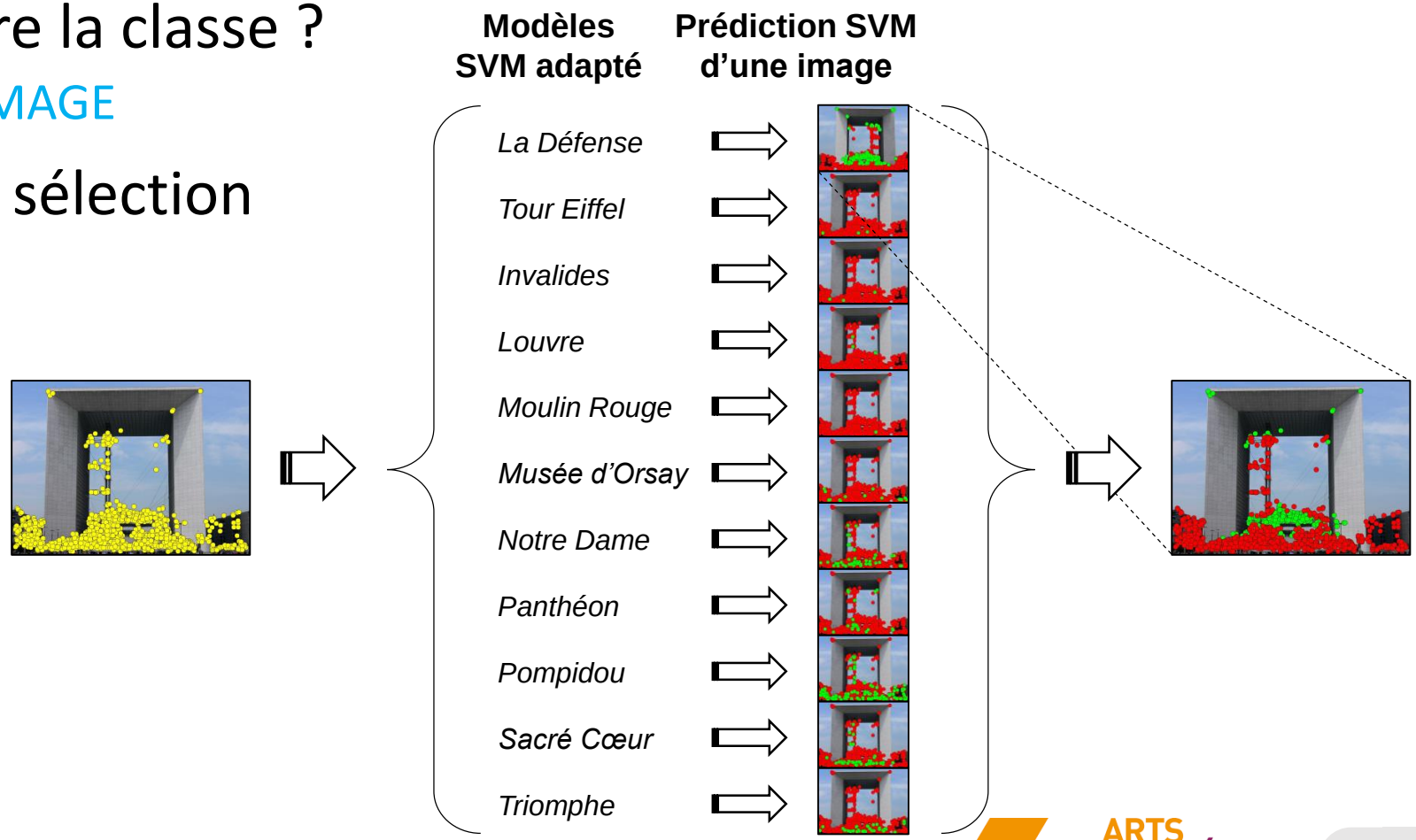
Modèles SVM adaptés par classe de bâtiments

- A quelle échelle prédire la classe ?
  STRATÉGIE LOCALE PAR DESCRIPTEUR
  - Sélection du classifieur retournant le meilleur score de confiance $P(X)$
- PAS DE COHÉRENCE GLOBALE



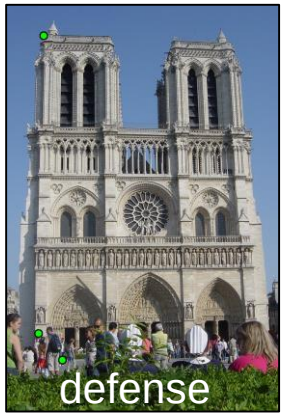
Modèles SVM adaptés par classe de bâtiments

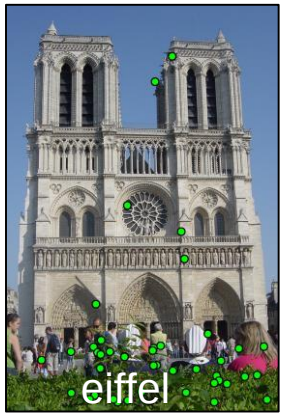
- A quelle échelle prédire la classe ?
  STRATÉGIE GLOBALE PAR IMAGE
- Besoin d'un CRITÈRE de sélection

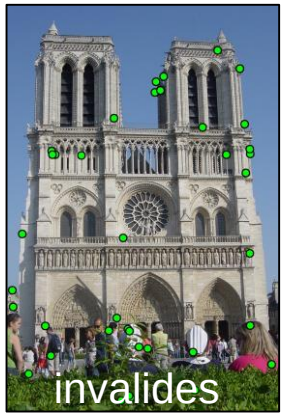


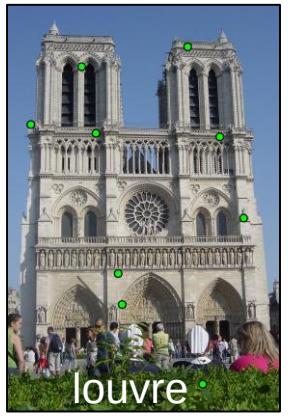
Modèles SVM adaptés par classe de bâtiments

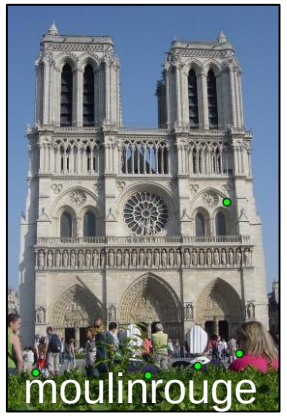
- #1 Critère de sélection du modèle SVM : NOMBRE de points *bâtiment*

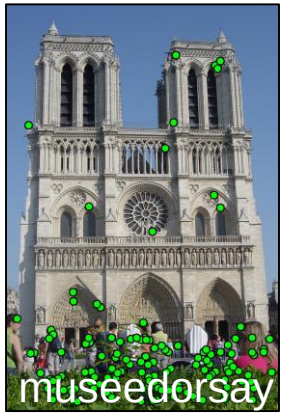


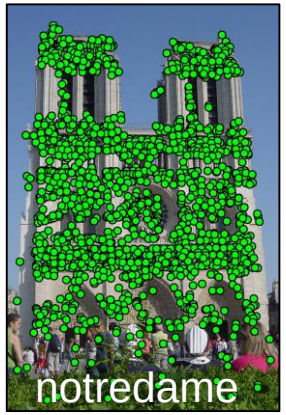


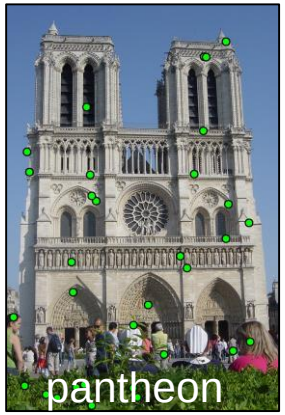


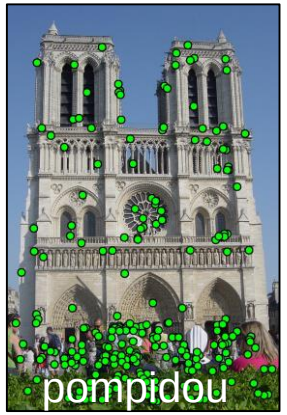


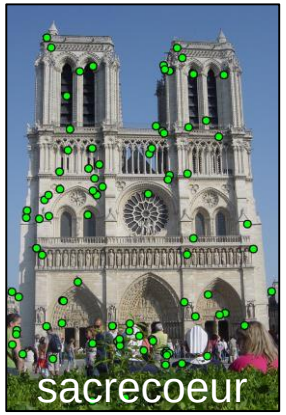


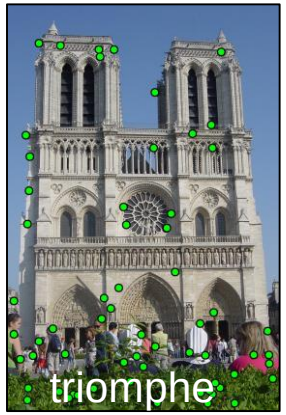












Modèles SVM adaptés par classe de bâtiments

- #1 Critère de sélection du modèle SVM : NOMBRE de points *bâtiment*

Paris6k		Oxford5k	
Classifieur	Descripteurs retenus après filtrage	Classifieur	Descripteurs retenus après filtrage
La Défense	6	All Souls	128
Tour Eiffel	40	Ashmolean	65
Invalides	38	Balliol	62
Louvre	55	Bodleian	77
Moulin Rouge	21	Christ Church	26
Musée d'Orsay	169	Cornmarket	167
Notre Dame	129	Hertford	132
Panthéon	53	Keble	234
Pompidou	368	Magdalen	39
Sacré Cœur	88	Pitt Rivers	35
Arc de Triomphe	82	Radcliffe Camera	135
Moyenne	**95**	**Moyenne**	**100**

Modèles SVM adaptés par classe de bâtiments

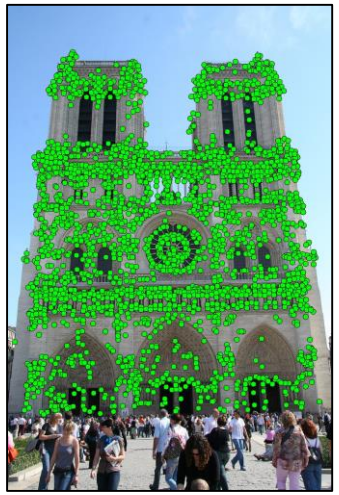
- #2 Critère de sélection du modèle SVM :
  SCORE DE CONFIANCE
  moyen de l'image

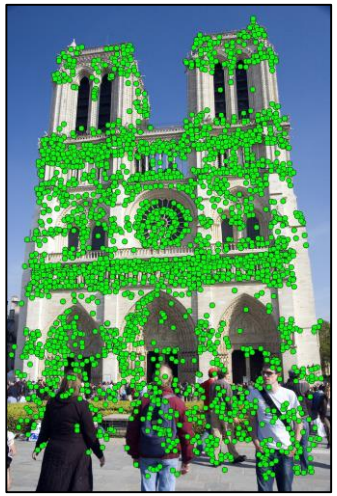
$$P(\mathrm{I}) = \frac{\sum_{X \in bâtiment} P(X)}{|X \in bâtiment|}$$

Paris6k		Oxford5k	
Classifieur	**Score de confiance**	**Classifieur**	**Score de confiance**
La Défense	0,526	All Souls	0,570
Tour Eiffel	0,541	Ashmolean	0,560
Invalides	0,542	Balliol	0,562
Louvre	0,531	Bodleian	0,569
Moulin Rouge	0,539	Christ Church	0,552
Musée d'Orsay	0,570	Cornmarket	0,571
Notre Dame	0,524	Hertford	0,569
Panthéon	0,560	Keble	0,577
Pompidou	0,453	Magdalen	0,556
Sacré Cœur	0,572	Pitt Rivers	0,553
Arc de Triomphe	0,561	Radcliffe Camera	0,565
Moyenne	**0,538**	**Moyenne**	**0,564**
Écart type	**0,031**	**Écart type**	**0,008**

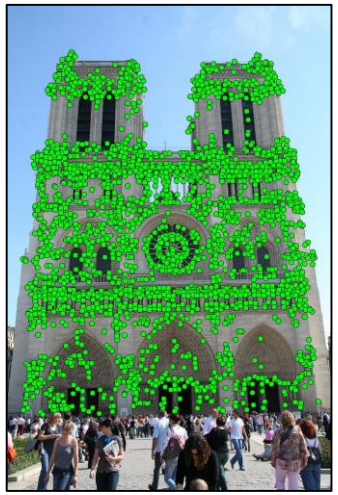
Modèles SVM adaptés par classe de bâtiments

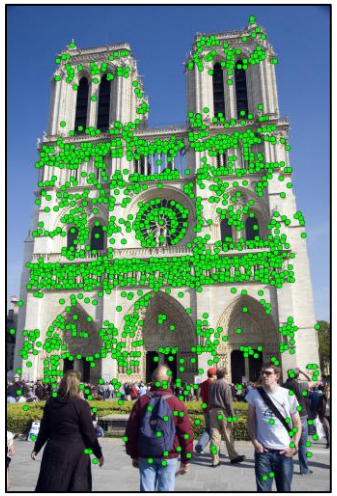
Classification globale



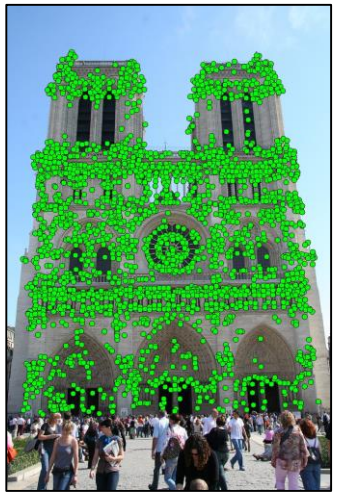


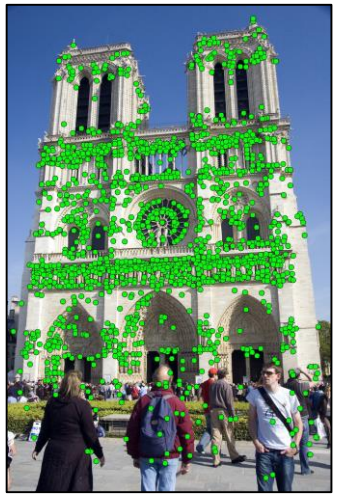
Classification adaptative avec nombre de points





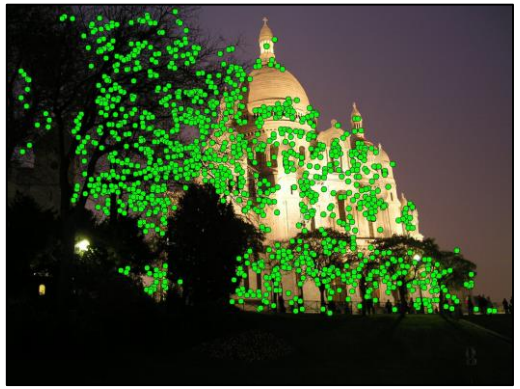
Classification adaptative avec score de confiance



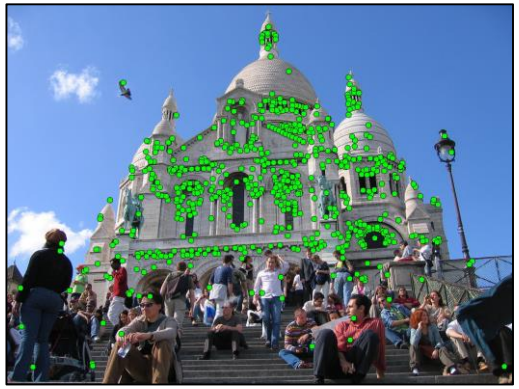


Modèles SVM adaptés par classe de bâtiments

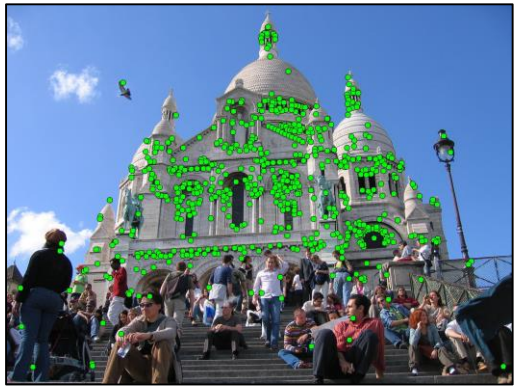
Classification globale



Classification adaptative avec nombre de points

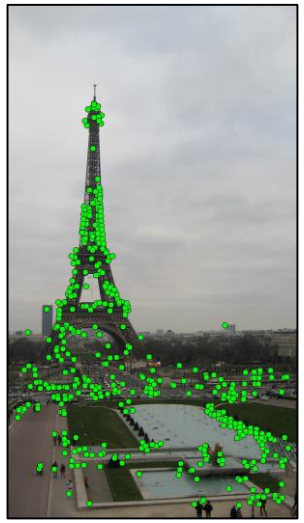


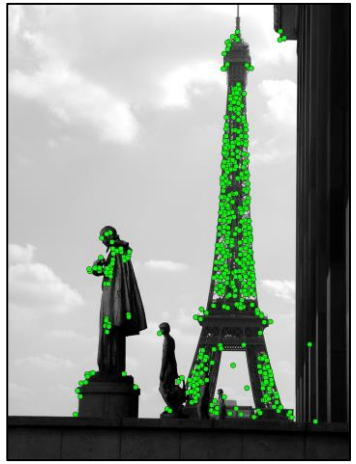
Classification adaptative avec score de confiance



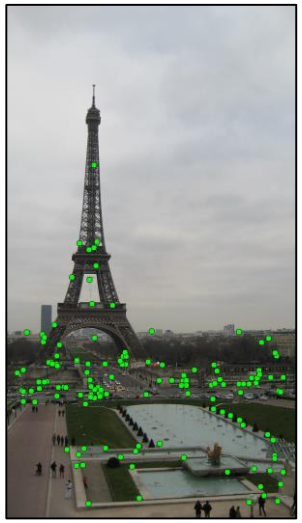
Modèles SVM adaptés par classe de bâtiments

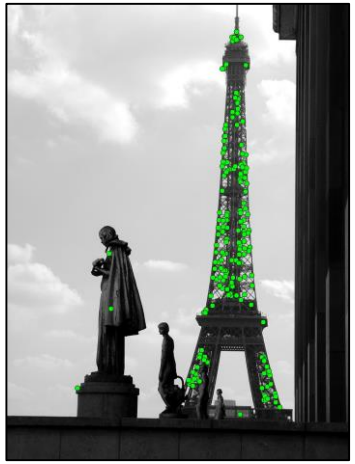
Classification globale



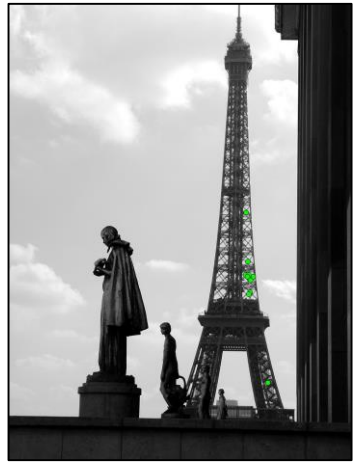


Classification adaptative avec nombre de points





Classification adaptative avec score de confiance

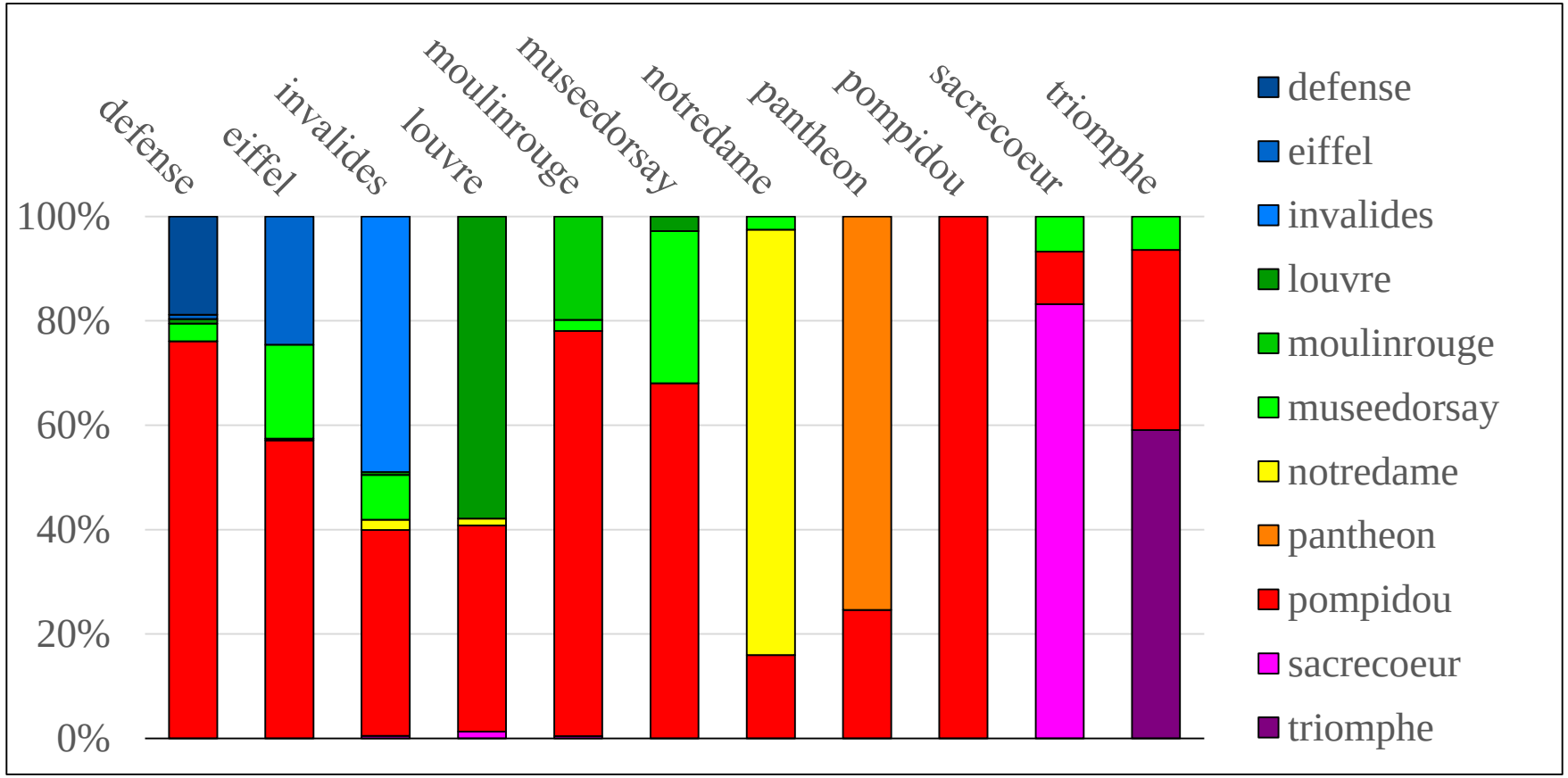


Modèles SVM adaptés par classe de bâtiments

Classification adaptée (fusion par score de confiance)				Classification globale		
Catégories	Faux positifs par image	Descripteurs bâtiment par image	Ratio	Faux positifs par image	Descripteurs bâtiment par image	Ratio
La Défense	1,660	93,855	0,018	372,950	612,100	0,609
Tour Eiffel	15,240	166,647	0,091	425,050	1040,950	0,408
Invalides	16,100	235,869	0,068	457,950	1027,150	0,446
Louvre	5,500	467,868	0,012	560,250	1227,200	0,457
Moulin Rouge	9,700	157,354	0,062	517,850	1066,400	0,486
Musée d'Orsay	139,660	594,306	0,235	1198,400	1768,900	0,677
Notre Dame	45,360	992,118	0,046	249,850	2013,350	0,124
Panthéon	8,140	440,794	0,018	200,100	1114,550	0,180
Pompidou	90,560	2593,882	0,035	383,400	3025,300	0,127
Sacré Cœur	22,380	576,483	0,039	285,450	1243,400	0,230
Arc de Triomphe	30,440	400,637	0,076	464,450	1295,250	0,359
Moyenne globale	**34,976**	**610,892**	**0,064**	**465,064**	**1403,141**	**0,373**

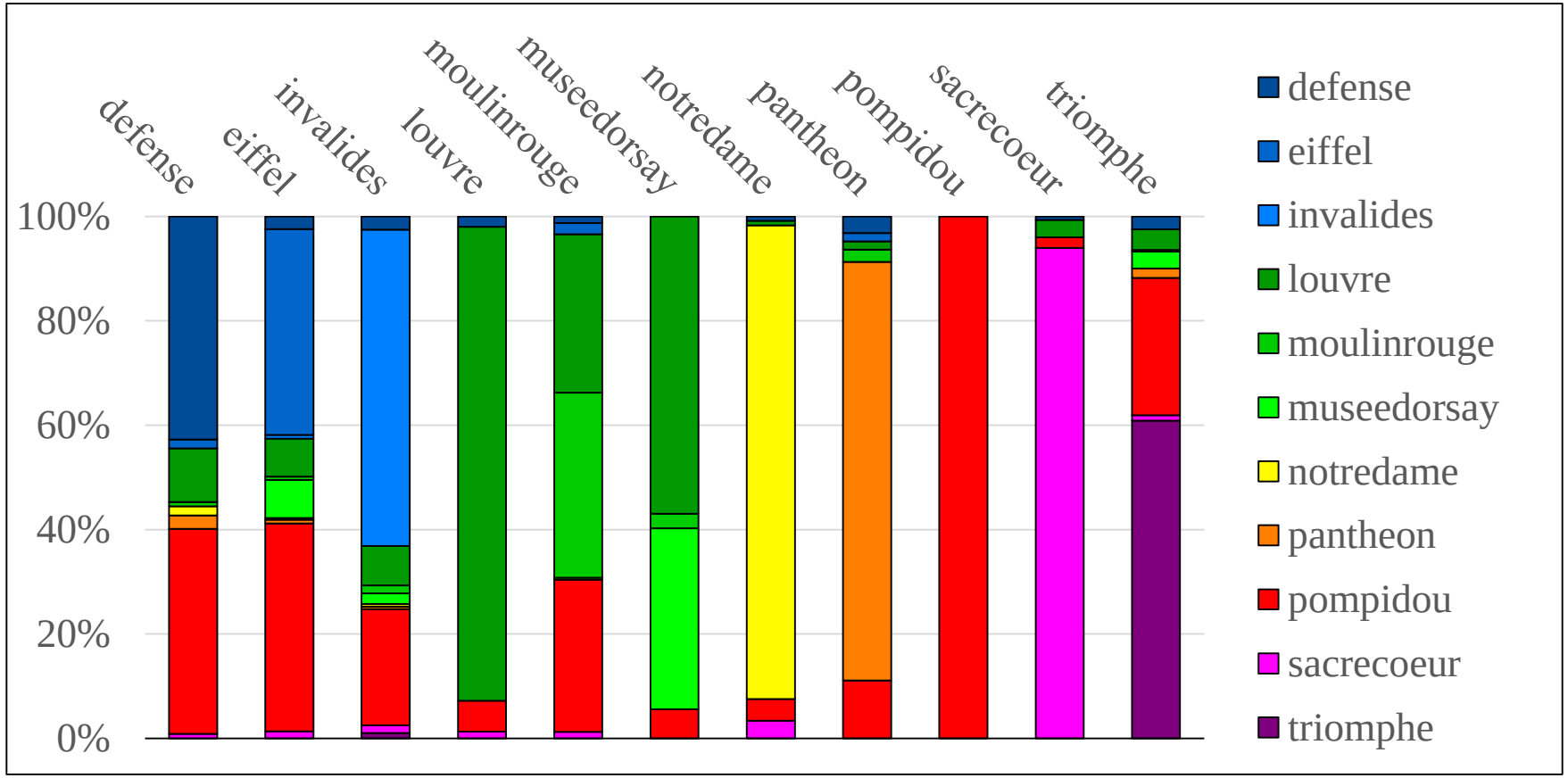
Modèles SVM adaptés par classe de bâtiments

- Choix du classifieur optimal (nombre de points *bâtiment*) Paris 6k



Modèles SVM adaptés par classe de bâtiments

- Choix du classifieur optimal (score de confiance)



Modèles SVM adaptés par classe de bâtiments

- Prédiction à l'échelle GLOBALE d'une image (plutôt que locale)
- Deux CRITÈRES DE SÉLECTION du modèle adapté à la catégorie de bâtiment de l'image (inconnue *a priori*)
  - NOMBRE de descripteurs prédits *bâtiment*
  - SCORE de confiance moyen de l'image
- Critère plus pertinent : SCORE DE CONFIANCE MOYEN DE L'IMAGE
  - Permet de s'affranchir de la variabilité importante du nombre de descripteurs extraits initialement entre les différentes catégories
  - Choix correct du modèle ⇔ Prédiction correcte de la CATÉGORIE de bâtiment

Critère	Paris6k	Oxford5k
nombre	54,41%	87,82%
confiance	66,32%	87,49%

Résultats de pourcentage de sélection correct du modèle SVM sur les images de la vérité terrain

Plan

- Introduction
- Etat de l’art et contributions
- Bases de données d’expérimentation
- Classification globale des descripteurs locaux
- Modèles SVM adaptés par classe de bâtiments
- Vérification et correction géométrique
- Recherche par similarité : résultats expérimentaux
- Conclusion

Vérification et correction géométrique

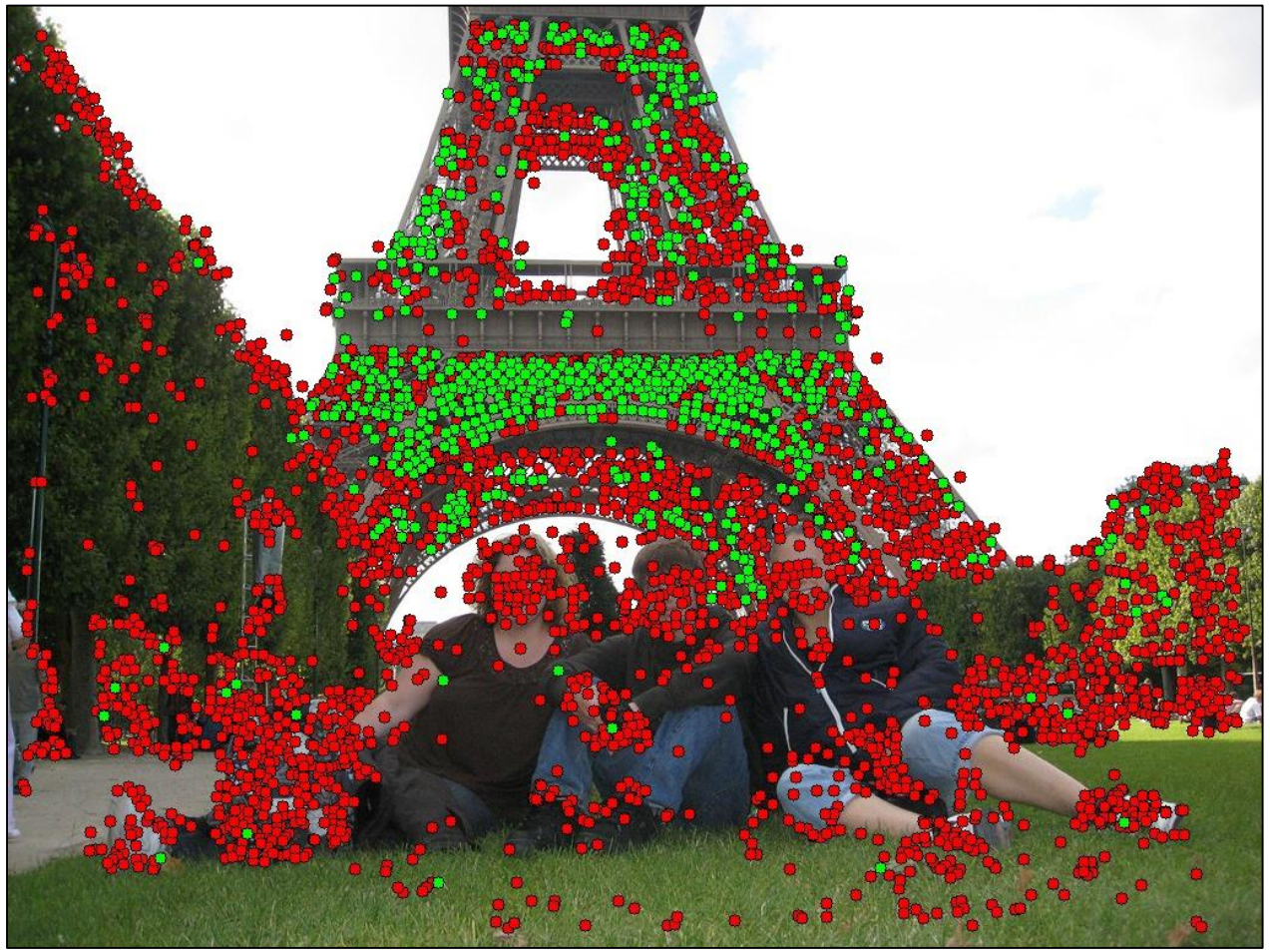
- Certains points clefs sont attribués à la classe *bâtiment*

Vérification et correction géométrique

- Certains points clefs sont attribués à la classe *bâtiment*
- Alors qu'ils devraient être attribués à la classe *non-bâtiment*

Vérification et correction géométrique

- Certains points clefs sont attribués à la classe *bâtiment*
- Alors qu'ils devraient être attribués à la classe *non-bâtiment*
- Lorsque leurs voisins sont effectivement de la classe *non-bâtiment*

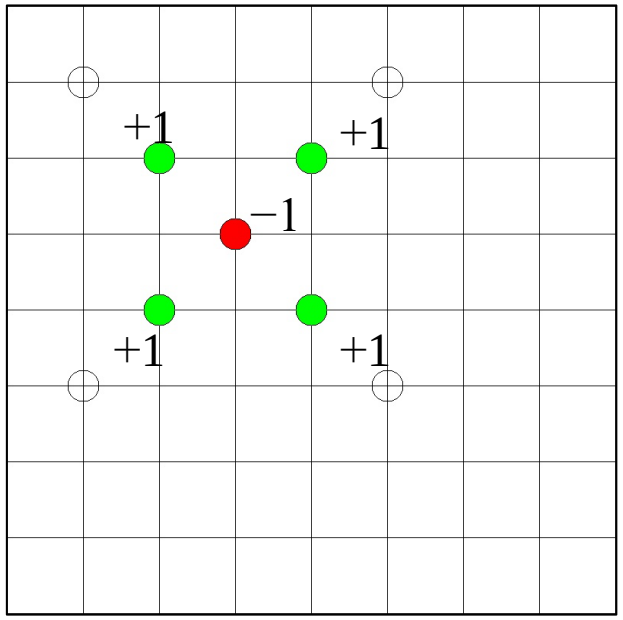


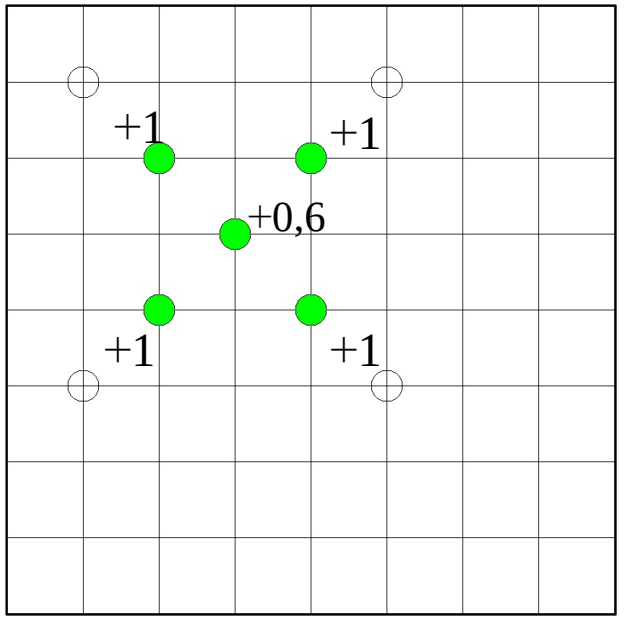
Vérification et correction géométrique

- Définition de la vraisemblance moyenne dans un voisinage

$$V(X) = \frac{\sum_{X \in Voisinage(p)} P[X] \cdot \Delta_X}{\sum_{X \in Voisinage(p)} P[X]}$$

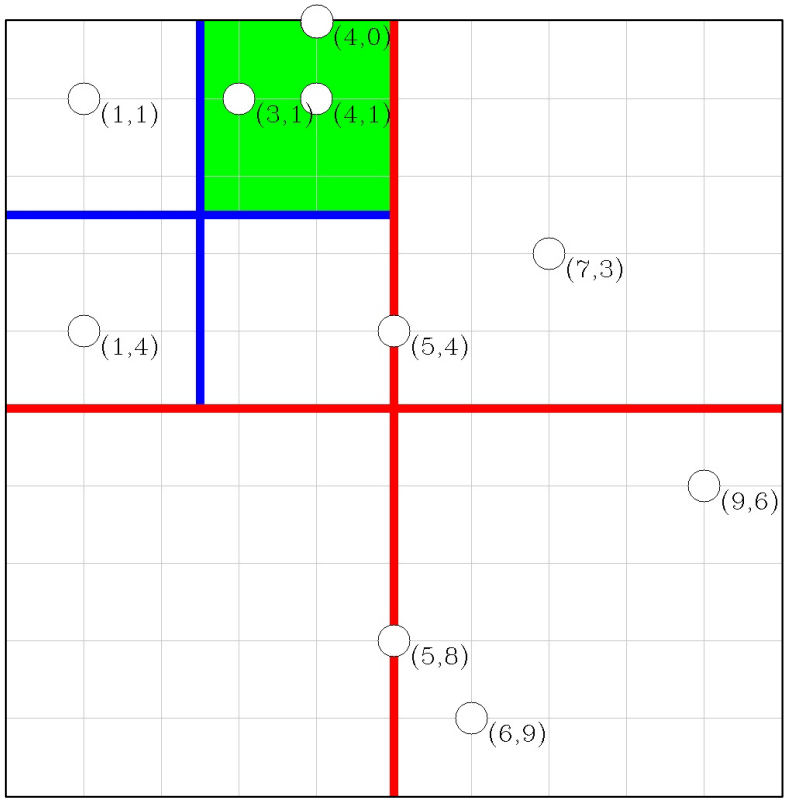
- Réattribution d'une valeur signée suivant l'influence du voisinage de X

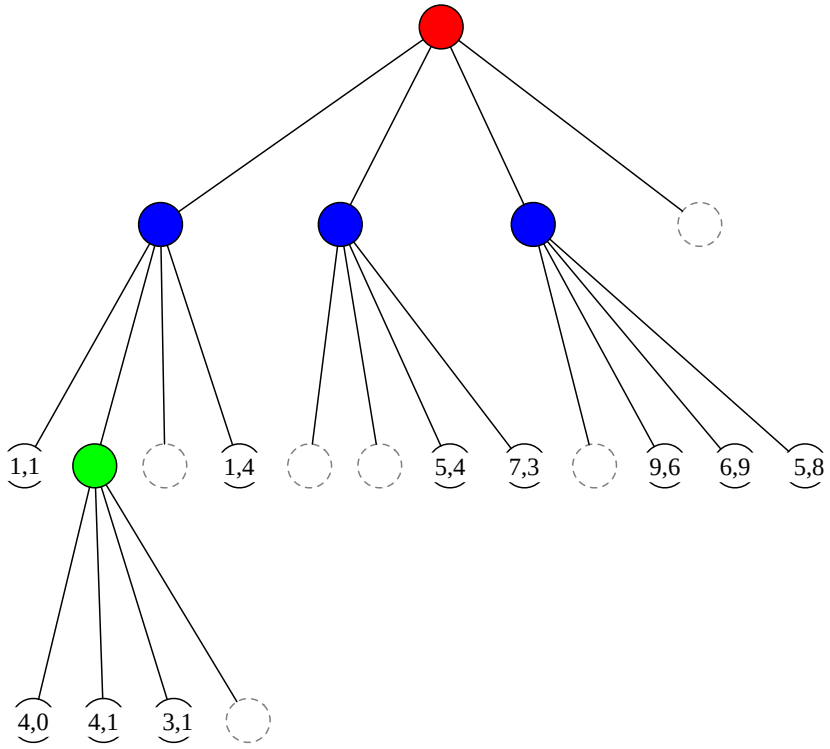




Vérification et correction géométrique

- Définition du voisinage : arbre de recherche quadtree

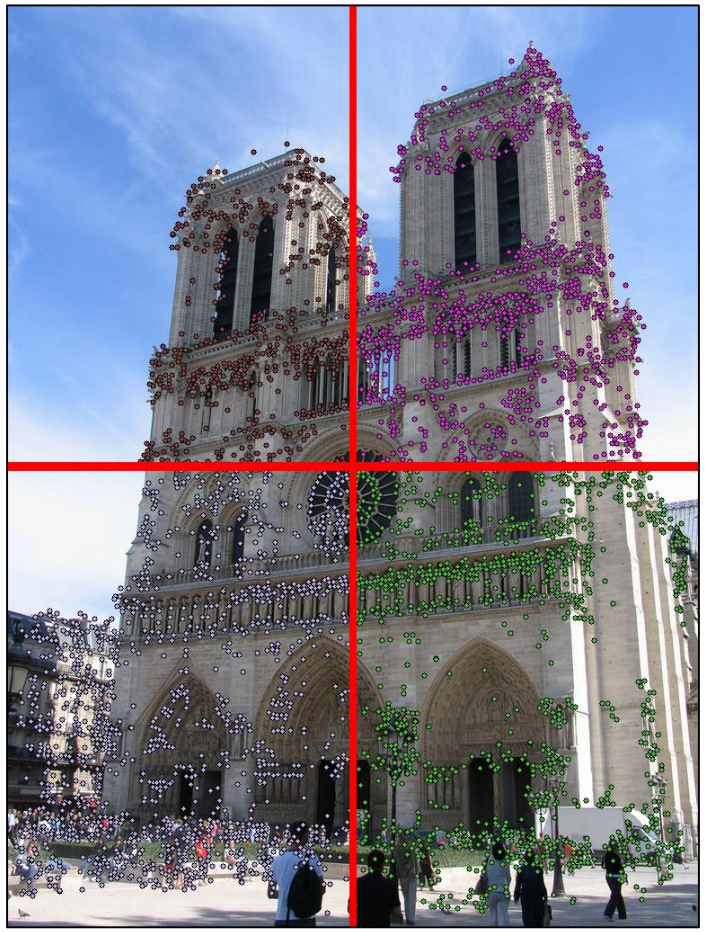




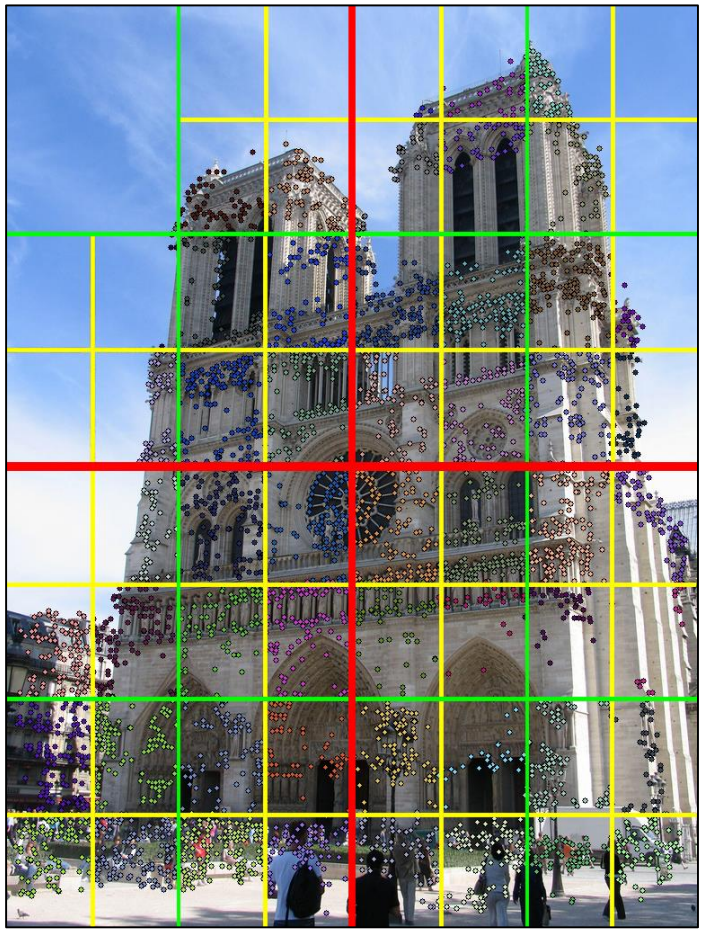
Vérification et correction géométrique

- Définition du voisinage

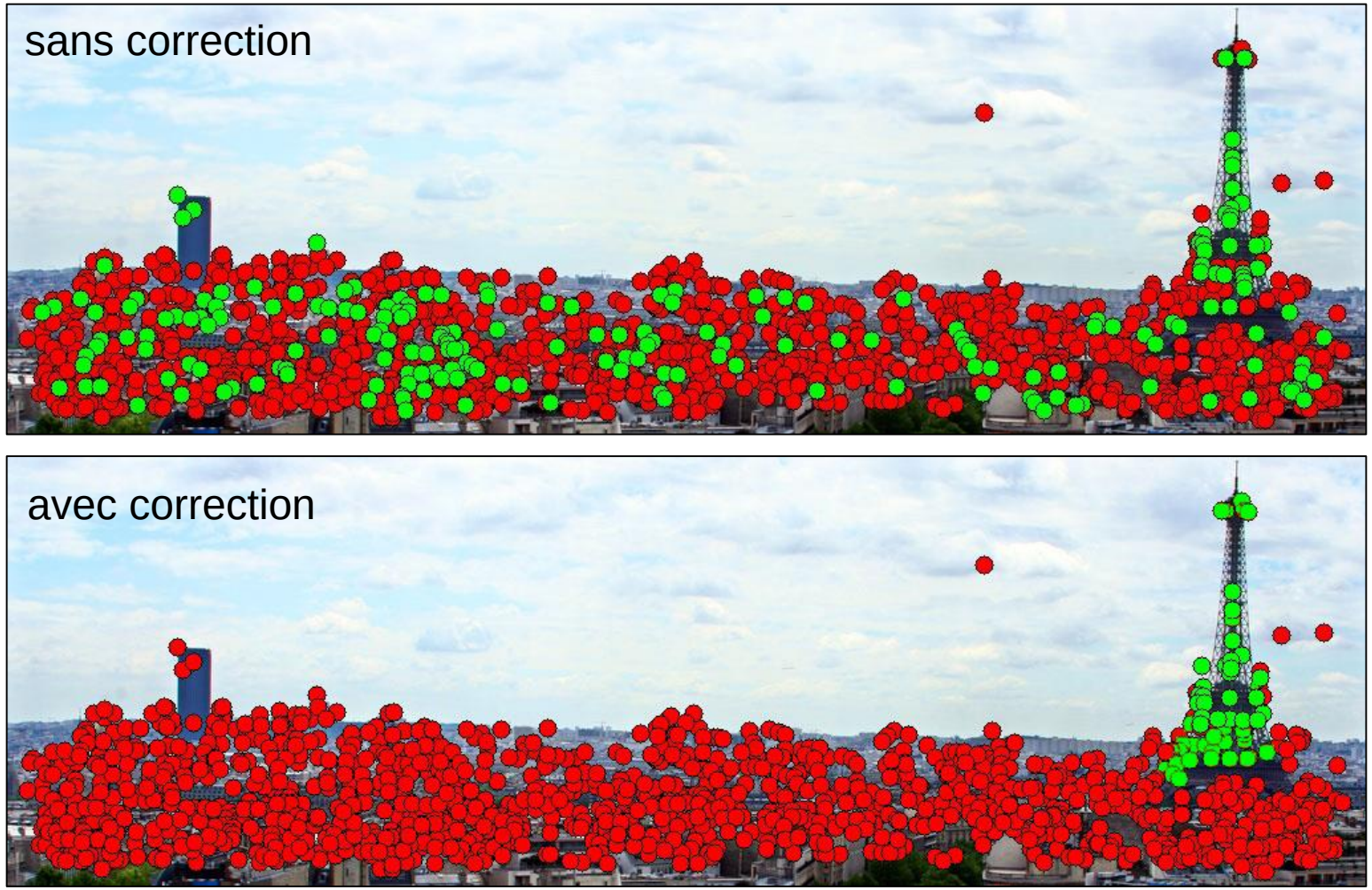
Profondeur = 1



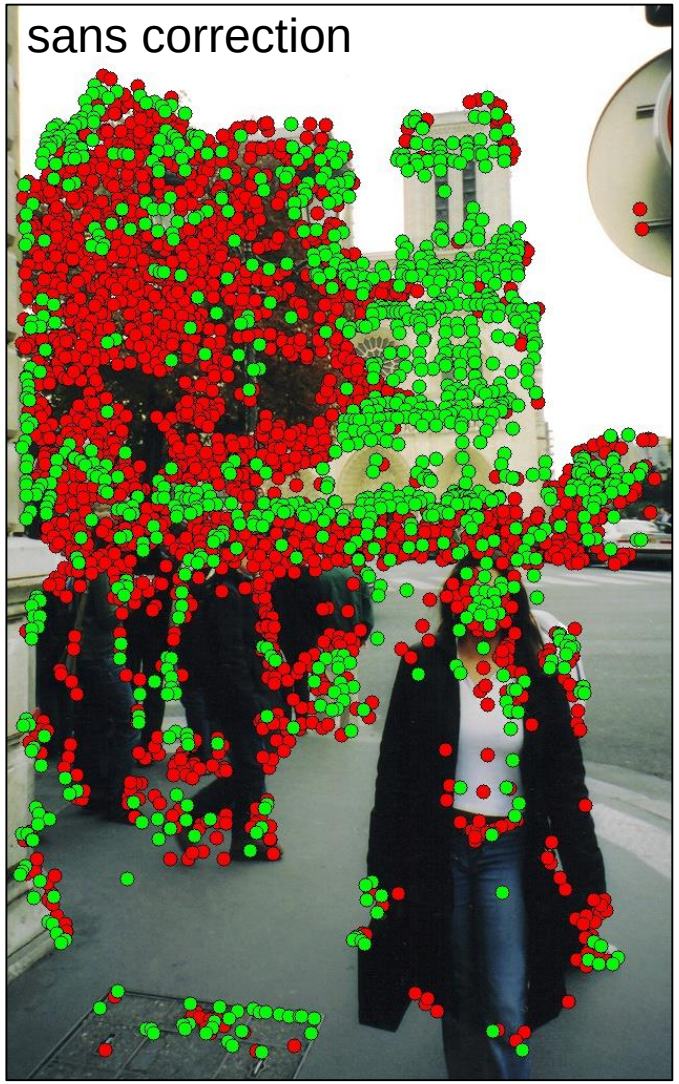
Profondeur = 3

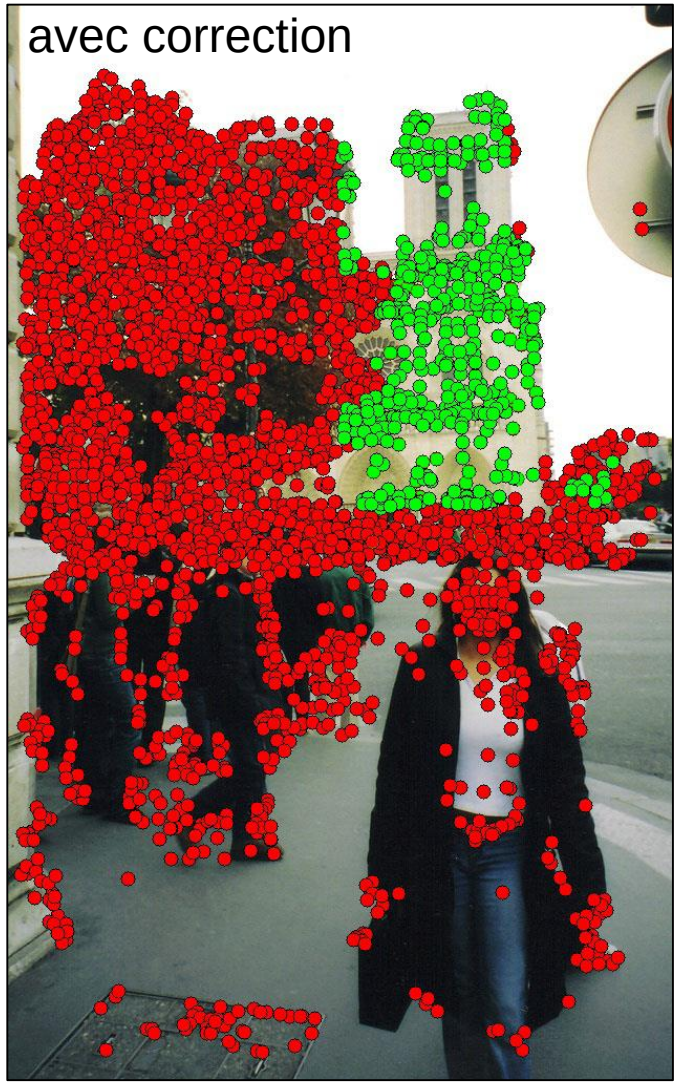


Résultats avec correction géométrique

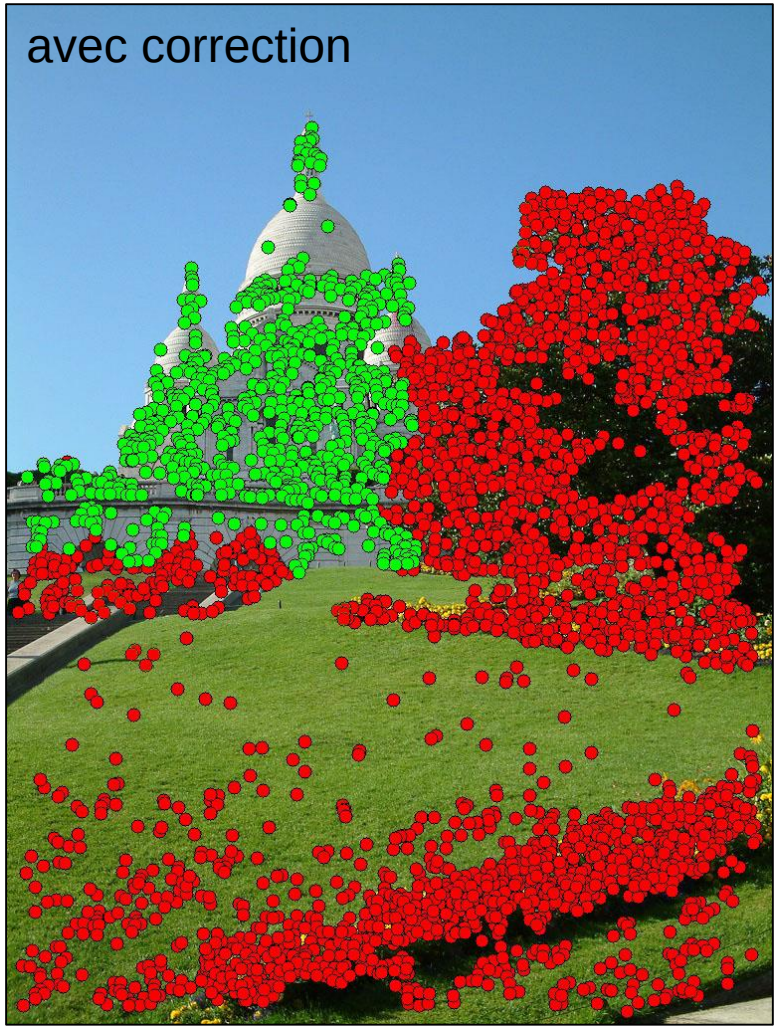


Résultats avec correction géométrique





Résultats avec correction géométrique



Vérification et correction géométrique

- Diminution du nombre de points d'intérêt

Nombre de descripteurs		Sans correction	Avec correction
Unique classifieur SVM global	Min	612	553
	Max	3025	3462
	Moyenne	1403	1449
	Écart-type	625	752
Classifieurs SVM adaptés choisis selon le nombre	Min	280	137
	Max	2374	2543
	Moyenne	690	537
	Écart-type	561	652
Classifieurs SVM adaptés choisis selon la confiance	Min	192	38
	Max	2374	2474
	Moyenne	612	422
	Écart-type	586	666

Vérification et correction géométrique

- Réduction significative du nombre de faux positifs

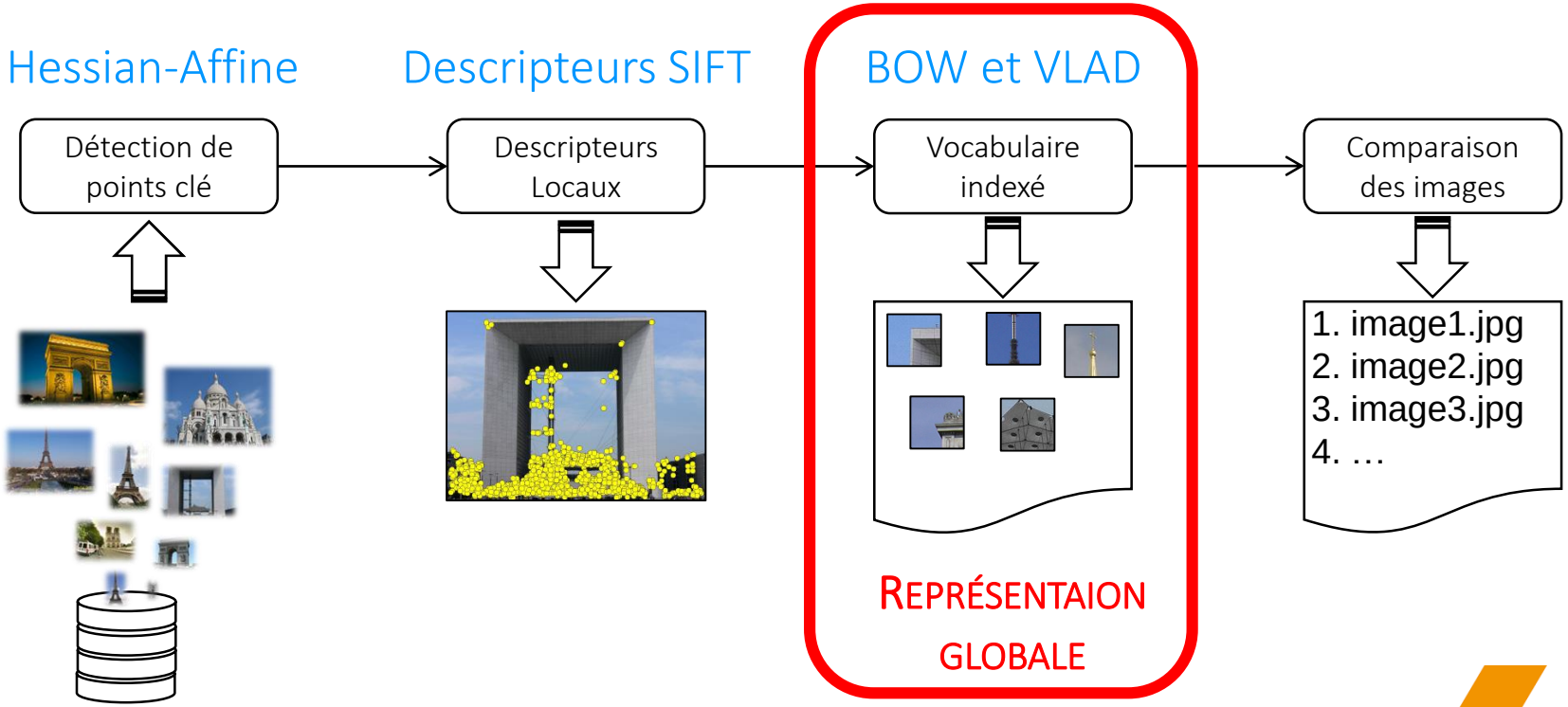
Nombre de descripteurs		Sans correction	Avec correction
Unique classifieur SVM global	Moyenne	0,373	0,315
	Écart-type	0,180	0,182
Classifieurs SVM adaptés choisis selon le nombre	Moyenne	0,323	0,177
	Écart-type	0,208	0,173
Classifieurs SVM adaptés choisis selon la confiance	Moyenne	0,210	0,164
	Écart-type	0,184	0,232

Plan

- Introduction
- Etat de l'art et contributions
- Bases de données d'expérimentation
- Classification globale des descripteurs locaux
- Modèles SVM adaptés par classe de bâtiments
- Vérification et correction géométrique
- Recherche par similarité : résultats expérimentaux
- Conclusion

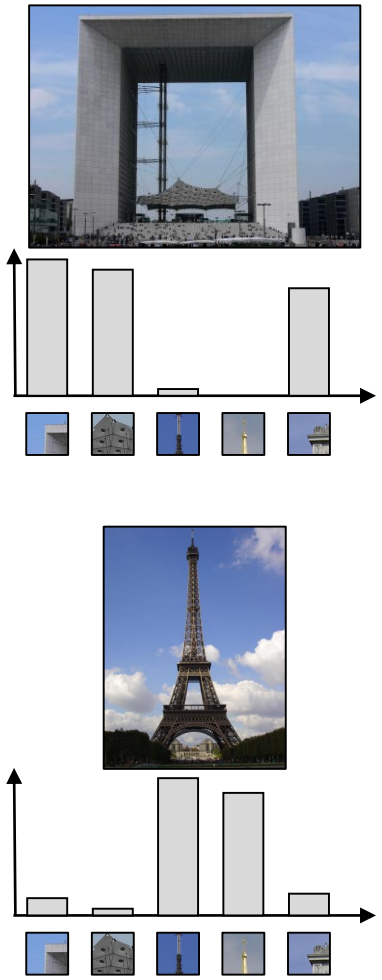
Globalisation de la représentation

- Recherche par SIMILARITÉ entre une image REQUÊTE et les image de la base de données
- Deux approches pour obtenir une représentation globale d'image
  - Modèle BOW (*Bag of Words*)
  - Modèle VLAD (*Vector of Locally Aggregated Descriptors*)

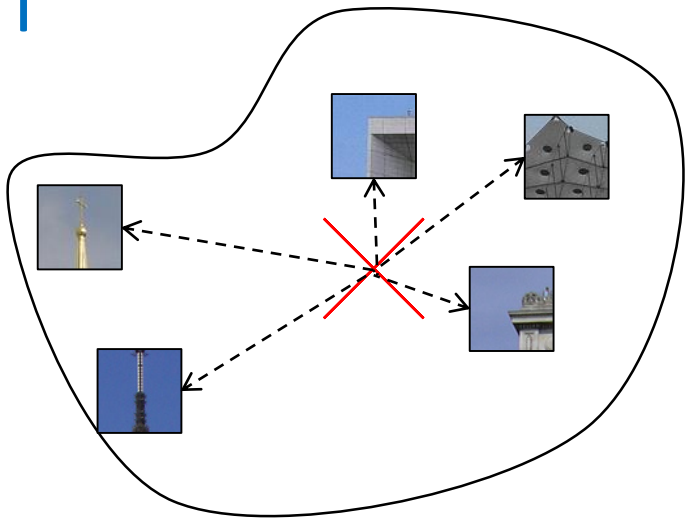


Globalisation de la représentation

- Modèle de BAG OF WORDS
  - Détermination d'un VOCABULAIRE de prototypes dans l'espace des descripteurs SIFT (CLUSTERING k-moyennes)
  - Construction d'un HISTOGRAMME DE FRÉQUENCE des MOTS VISUELS de la base de donnée pour chaque image
- Pondération TF-IDF (term frequency – inverse document frequency)
  - $\mathrm{tf}_{i,j}$: nombre de mot i dans le document j par rapport au total de mots dans le document j
  - idf_i : nombre total de document par rapport au nombre de document contenant le mot i
  - $\mathrm{tfidf}_{i,j} = \mathrm{tf}_{i,j} \cdot \mathrm{idf}_i$
  - Limite le poids des mots trop communs
  - Favorise les mots discriminants
- Mesure de SIMILARITÉ entre deux vecteurs u et v
  - Distance euclidienne : $\Delta(u,v) = \sqrt{\sum_{i=1}^{n}(u_i - v_i)^2}$
  - Distance cosinus : $\cos(u,v) = \frac{u \cdot v}{\|u\| \cdot \|v\|}$



Globalisation de la représentation



- Modèle VLAD (*Vector of Locally Aggregated Descriptors*)
  - Vocabulaire de caractéristiques locales (SIFT)
  - Agrégation de vecteurs résiduels
  - Différence entre un descripteur et le centre du cluster
  - N descripteurs locaux SIFT $X = \{x_1, \dots, x_n\}$
  - Vocabulaire $C = \{c_1, \dots, c_k\}$ (n descripteurs, en k clusters)

$$x \mapsto n(x) \arg\min_{c \in C} \|x - c\|^2$$

centre du cluster le plus proche attribué au descripteur local x

- Vecteur résiduel

$$v_i = \sum_{j \in [1,n]} x_j - c_i$$

- VLAD

$$\boldsymbol{VLAD} = [v_i \dots v_k]$$

Résultats expérimentaux

- Mesures objectives d'évaluation

$$rappel = \frac{|Correctes\ renvoyées|}{|Correctes\ selon\ la\ vérité\ terrain|}$$

$$précision = \frac{|Correctes\ renvoyées|}{|Total\ renvoyées|}$$

$$F = 2\ \times \frac{pécision\ \times\ rappel}{précision + rappel}$$

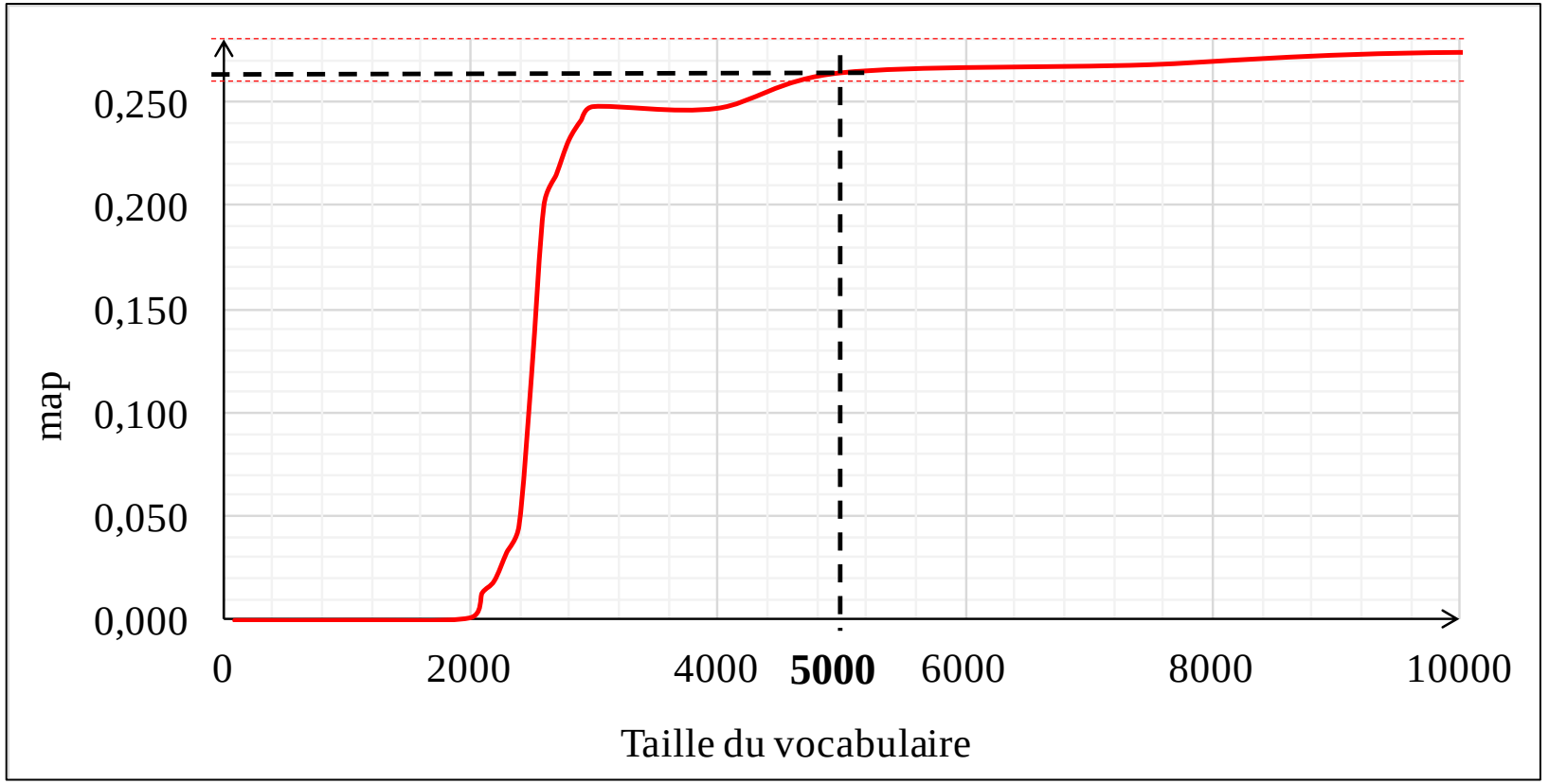
$$MAP = \frac{\sum_{k=1}^{Q} AP(k)}{Q}$$

$$AP(n) = \frac{\sum_{k=1}^{n} P(k) \times \delta(k)}{|Correctes|}$$

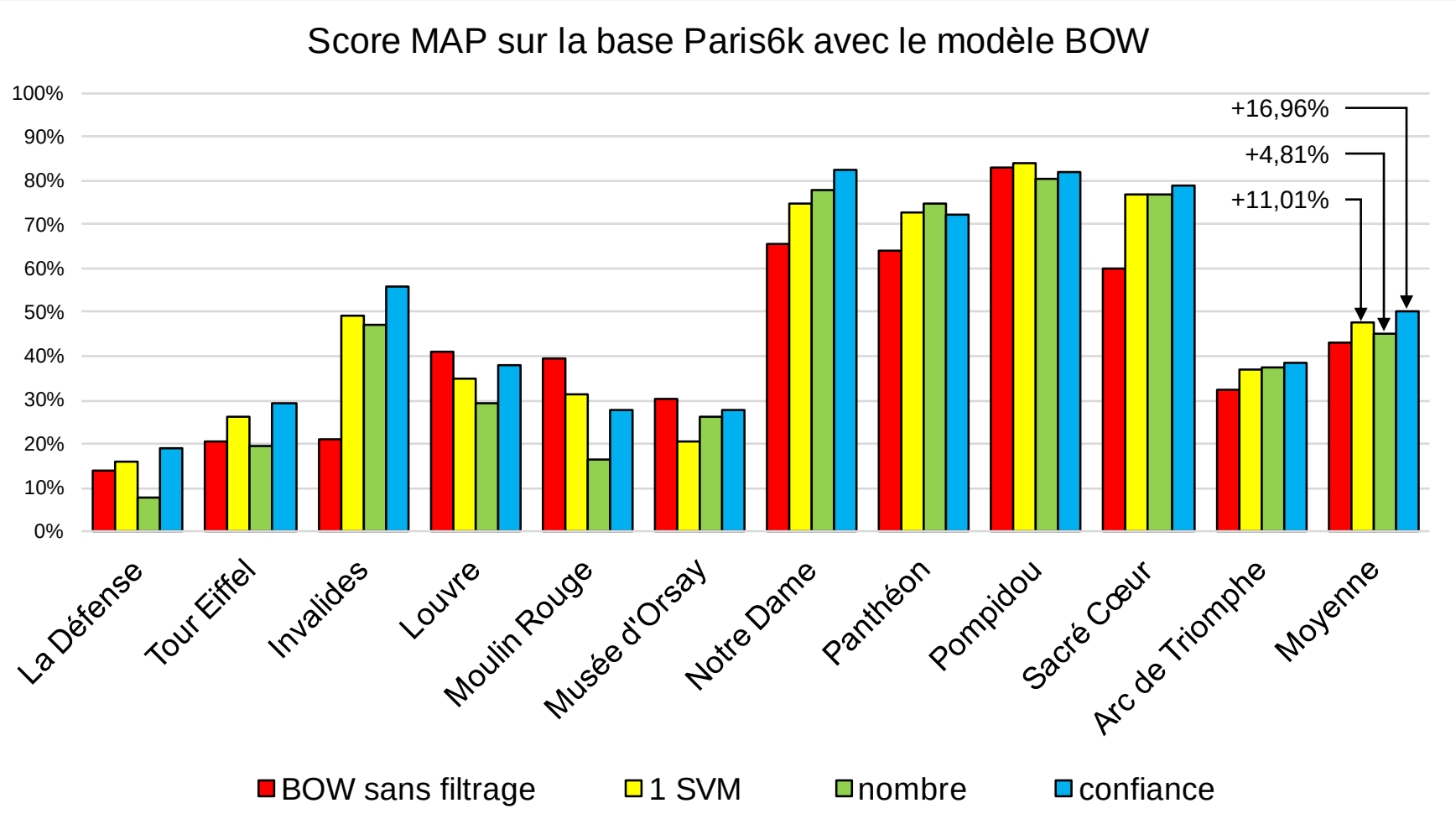
$\delta(k)$: fonction d'indication valant 1 si l'image retournée au rang k est correcte et 0 sinon

Résultats expérimentaux

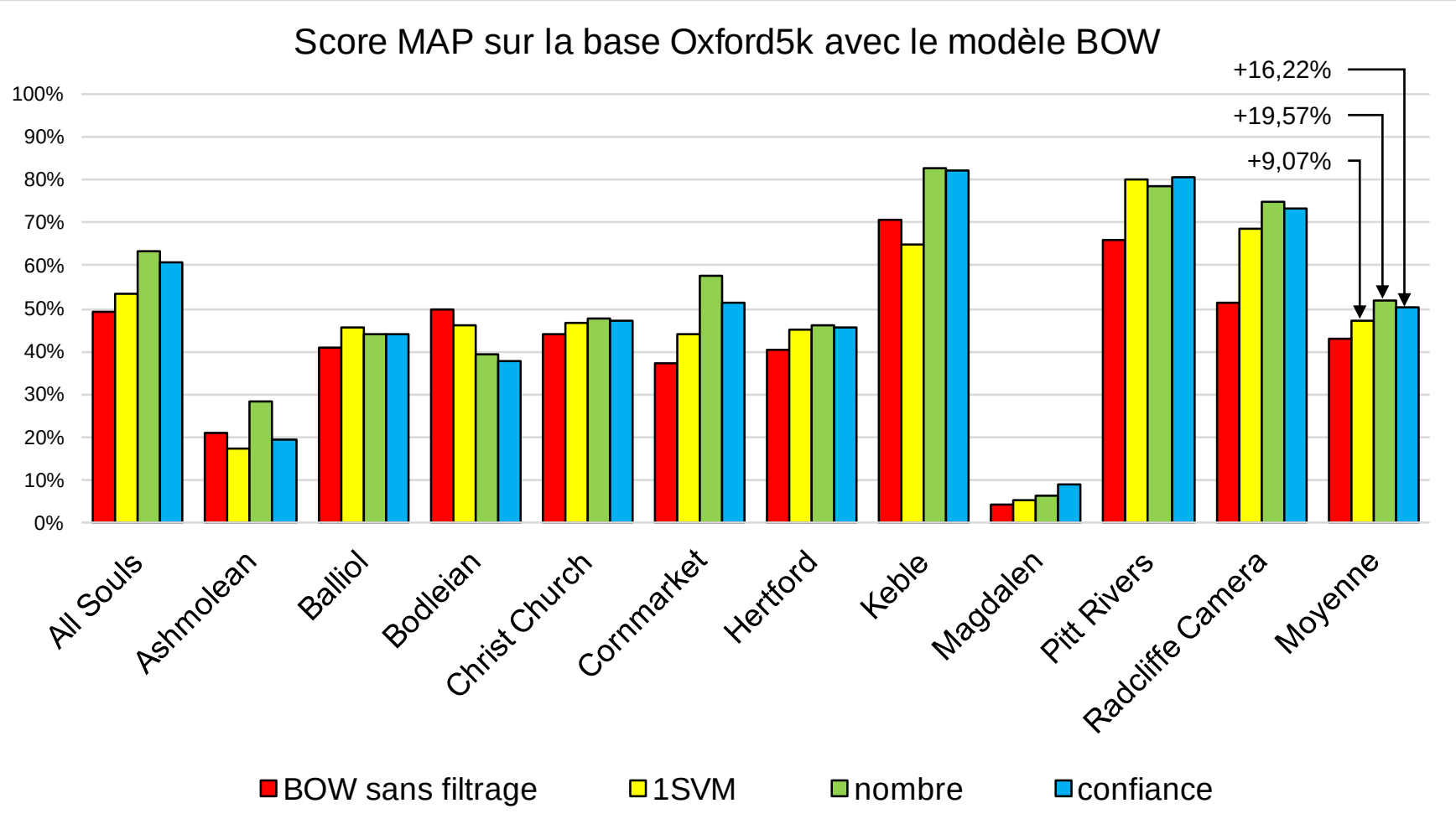
- Paramètres du *Bag of Words*
  - Évolution du score MAP en fonction de la taille du vocabulaire sur le corpus Paris 6k



Résultats expérimentaux



Résultats expérimentaux



Résultats expérimentaux

- Scores MAP avec le modèle VLAD
  - Le filtrage des points clés pour certaines images n'en retient qu'une dizaine
  - Risque d'erreurs de reconnaissance dues à un manque de données
  - Nouvel ensemble d'images rejetant les images avec moins de 10 descripteurs filtrés

Base de données réduite	Paris6k	Oxford5k
Sans filtrage	6 412	5 063
1 SVM	6 381	5 057
nombre	6 363	5 033
confiance	5 735	4 843

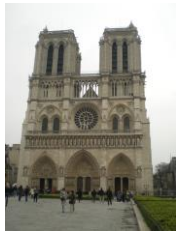
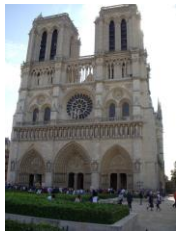
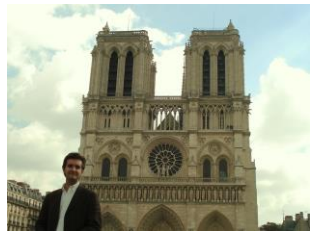
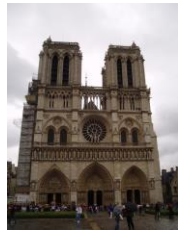
Scores MAP	Paris6k	Oxford5k
VLAD sans filtrage	0,506	0,537
VLAD / nombre	0,527	0,549
VLAD / confiance	0,540	N/A

Scores MAP	Paris6k	Oxford5k
BOW sans filtrage	0,428	0,432
BOW / nombre	0,449	0,517
BOW / confiance	0,501	0,502

Résultats expérimentaux

- Exemple 1

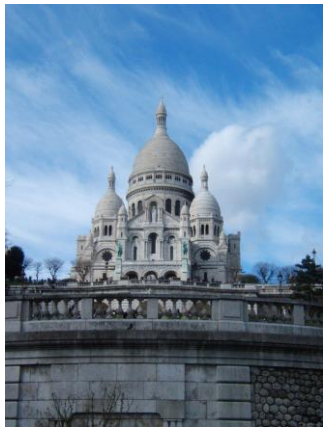
Image requête

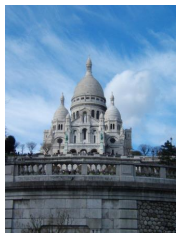
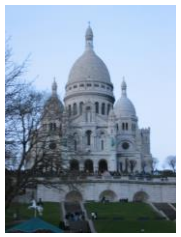
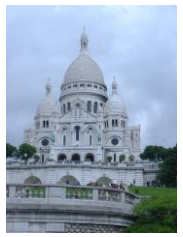
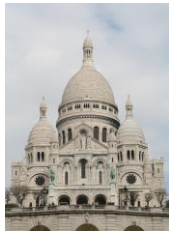
1. 
5. 
9. 
2. 
6. 
10. 
3. 
7. 
11. 
4. 
8. 
12. 

Résultats expérimentaux

- Exemple 2

Image requête



1. 
5. 
9. 
2. 
6. 
10. 
3. 
7. 
11. 
4. 
8. 
12. 

Plan

- Introduction
- Etat de l’art et contributions
- Bases de données d’expérimentation
- Classification globale des descripteurs locaux
- Modèles SVM adaptés par classe de bâtiments
- Vérification et correction géométrique
- Recherche par similarité : résultats expérimentaux
- Conclusion

Conclusion

- Sélection PERTINENTE et SÉMANTIQUE de points d'intérêts d'un objet particulier dans une scène complexe
- CLASSIFICATION AUTOMATIQUE des descripteurs par différents modèles SVM adaptés
- Définition de DEUX MESURES pour le CHOIX D'UN MODÈLE SVM optimal dans une catégorie de bâtiment donnée
- INTERPRÉTATION des objets présents dans une scène complexe urbaine
- ROBUSTESSE testée sur différentes bases (Paris 6k et Oxford 5k) de données et avec deux modèles de représentation (BOW et VLAD)

Imagerie et reconnaissance d'objet