



Le 1er Mars 2018

Junshuai ZHU
Yihong XU

Sujet 6 : Descripteurs locaux pour la recherche d'images similaires

TP Lot1

Ingénieur Généraliste
F4B 516
Encadrant : John Puentes



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

1. PROBLEMATIQUE

Le but de ce projet est de rechercher de manière automatique les objets similaires dans les images dont l'angle de visualisation et la condition lumineuse changent. L'entrée est une image ayant un objet principal et la sortie souhaitée est des images des objets les plus similaires. Les descripteurs utilisés sont ORB, Brief et SIFT.

2. EXTRACTION DES POINTS CARACTERISTIQUES

2.1 Résultat de l'extraction des points caractéristiques

Les nombres de points caractéristiques détectés sont très variés (SIFT 178 points, Brief 75 points et ORB 478 points). En plus, nous constatons que les points détectés par Brief (*Star detector*) sont plus dispersés de l'objet par rapport aux autres détecteurs (DoG pour SIFT ; détecteur rapide pour ORB). Une remarque est que la taille de l'image a une influence importante au nombre de points caractéristiques détectés. Une taille de pixels assez grande est nécessaire pour la détection des *keypoints* dans l'objet. Dans ce projet, nous avons choisi une taille de 384x288.

3. INTRODUCTION DE LA BASE DE DONNEES

Nous utilisons une partie de la base d'images ALOI (Amsterdam Library of Object Images) et 2 groupes d'images ont été choisis :

Changement de température de couleur (100 classes x 12/classe)

Changement de l'angle de visualisation (100 classes x 3/classe)

Pour chaque classe (type) d'objet, il y a, au total, 15 images. Nous avons choisi 100 classes pour ce projet et cela fait 1500 (15x100) images. Ces images ont été divisées en deux parties. 2/3 sont attribuées au groupe de l'entraînement (*clustering* par K-means) et servent comme la base de recherche. Le reste 1/3 sont la base de test. En résumé, pour chaque classe, 10 images (2 images de rotation + 8 images de changement de température) forment la base de recherche et 5 images sont pour le test (1 image de rotation + 4 images de température).

4. CHAINE DU TRAITEMENT

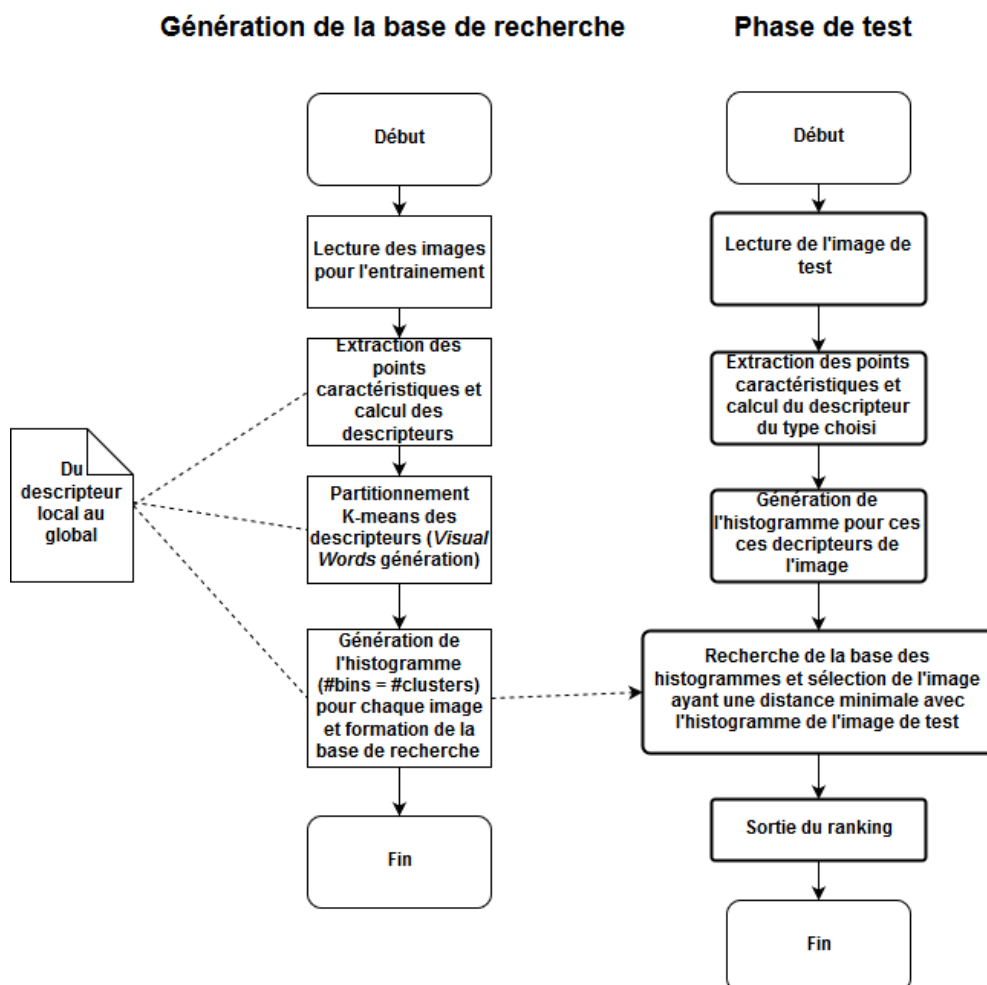


Figure 1 : Schéma du traitement.

5. PHASE DE TESTS ET COMPARAISONS

Premièrement, nous faisons la comparaison du graphe P-R qui est une métrique faisant la courbe P-R ($Recall = \frac{TP}{TP+FN}$, $Precision = \frac{TP}{TP+FP}$) pour les classeurs binaires. Pour chaque classe, toutes les images de test (5 exemples positifs) de cette classe sont sélectionnées et à cause de temps de calcul, seulement 10 autres images venant d'autres classes sont choisies, de manière aléatoire (exemples négatifs). A partir de ces 15 images, un graphe P-R est construit pour chaque classe et nous évaluons son AUC (Area Under Curve).

Le graphe P-R s'adapte au classifieur binaire, ce qui n'est le cas dans notre problème. En plus, P-R n'évalue que sur la probabilité ou le score du niveau de ressemblance entre l'image en entrée et toutes les images dans la base de la classe évaluée. Autrement dit, le niveau de ressemblance avec d'autres classes n'est pas évalué. Il n'est pas donc approprié naturellement dans notre cas (multi-classe et sortie étant la distance). Pour cela, nous avons adopté une approche *one v.s all* et le score se fait par l'équation ci-dessous :

$$score = \frac{\sum_{i \text{ toutes images de cette classe dans la base}} 1}{distance \text{ à l'image } i \text{ de cette classe dans la base} + 0.001}$$

De cette manière, l'image ayant une distance minimale à cette classe a le plus grand score.

5.1 Performance des différents descripteurs

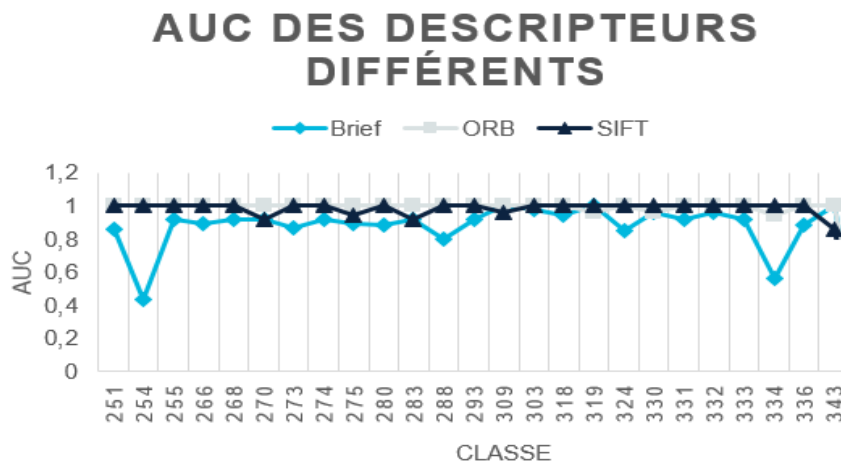


Figure 2 : AUC des descripteurs différents.

Nous ne montrons ici que 27 classes ayant un AUC moins de 1 dans certains types de descripteurs. Nous fixons le nombre de clusters à 50 et le nombre maximum de points caractéristiques détectable pour une image est fixé à 100. Parmi ces descripteurs, l'ORB a une performance meilleure avec la valeur moyenne de l'AUC égale à 0.9944 ; SIFT a aussi une bonne performance avec la valeur moyenne égale à 0.984 ; BRIEF a une valeur moyenne égale à 0.8844. La performance est fortement liée au nombre de points détectés par les méthodes différentes ainsi au positionnement (distribution) des points.

5.2 Performance pour nombre des points caractéristiques différent

Pour voir l'influence du nombre de points caractéristiques sur la performance de la recherche des images similaires. Comme l'exemple, nous prenons le descripteur ORB et fixons le nombre de clusters du K-means à 50.

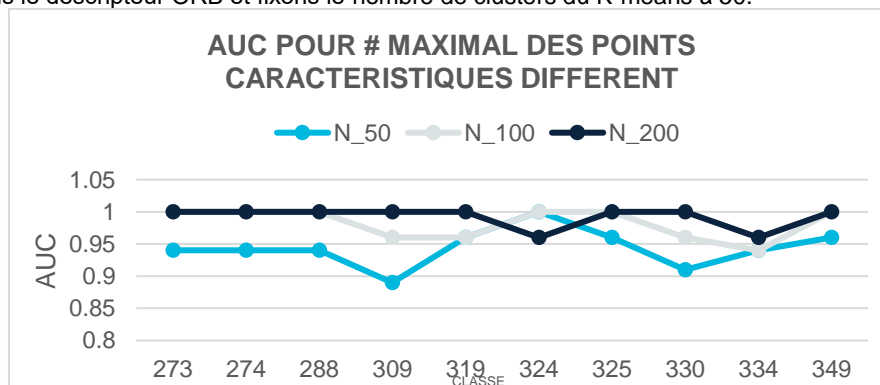


Figure 3 : AUC pour nombre des points caractéristiques différent.

Nous trouvons que la performance de la recherche pour des images similaires se dégrade en limitant le nombre maximum autorisé des points caractéristiques détectés. Cette dégradation de performance est raisonnable parce qu'un nombre de points très limité ne peut pas bien caractériser/représenter une image en entière. Par contre, N=100 et N= 200 n'ont pas de différence importante.

5.3 Performance pour nombre de clusters différent

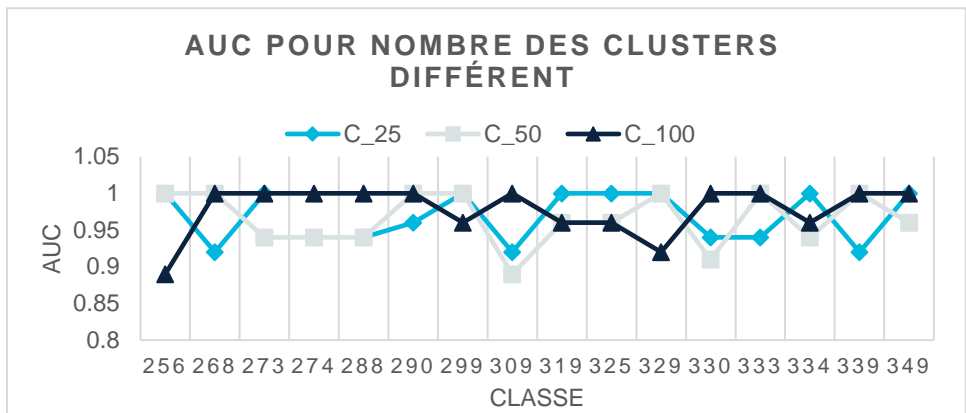


Figure 4 : AUC pour la quantité du cluster différente.

Dans le cas où le nombre de clusters est différente (nous testons sur l'ORB et le nombre maximum de points caractéristiques est fixé à 50), nous pouvons trouver que le nombre de clusters n'influence pas de manière conséquente sur la performance. La valeur moyenne de l'AUC pour la quantité de clusters égale à 25, 50 et 100 sont respectivement, 0.969, 0.965, et 0.978.

En revanche, il faut noter qu'un bon résultat en termes de valeur de l'AUC pour une classe ne veut pas dire que les images de même classe seront forcément sélectionnées lors de la recherche des images similaires. C'est parce qu'il existe parfois que même une image de classe i en entrée ayant une distance minimale à toutes les images de classe i par rapport aux exemples négatifs, il est probable qu'elle ait une distance plus petite à d'autres images d'une classe différente dans la base.

Pour cette raison, nous ensuite montrons ci-dessous le TOP10 Ranking en termes de distance avec les images dans la base et ceci reflète directement quelles images vont être sélectionnées lors de la recherche. Deux images de classe 251 ont été choisies (251_c.png et 251_i120.png) et #clusters = 50, #max keypoints = 100.

Ranking \ type	ORB		SIFT		Brief	
	251_c (rotation)	251_i120 (color température)	251_c (rotation)	251_i120 (color temperature)	251_c (rotation)	251_i120 (color temperature)
1	251_i190	251_i130	251_i210	251_i130	260_l	251_i210
2	251_i180	251_i110	251_i190	251_i110	304_i180	319_i210
3	251_i110	251_i180	251_i230	251_i230	304_i170	251_i250
4	251_i230	251_i140	251_i180	251_i180	309_c	319_i190
5	251_i250	251_i210	336_i170	251_i190	285_i110	319_c
6	251_i210	251_i250	251_i250	251_i140	285_i170	251_i180
7	251_i130	251_i230	278_c	265_i120	285_i180	319_i120
8	251_i140	251_i190	336_i110	313_i250	340_i130	251_i190
9	346_i110	260_l	336_i120	277_c	300_l	251_i110
10	346_i230	251_l	336_i190	300_l_	285_i190	251_i130

Tableau 1 : Ranking de la recherche. (xxx_i sont les images avec changement de température de couleur et le reste sont celles tournées)

D'une part, l'objet tourné est difficile à trouver ses images similaires parce que la forme est déformée et une partie de l'objet est cachée. En revanche, le changement simple de la luminosité n'est pas gênant pour la détection. D'autre part, nous pouvons voir que l'ORB a un meilleur résultat pour cette image (classe 251) et souvent l'ORB est capable de trouver plus de points caractéristiques que d'autres méthodes. La performance de Brief s'améliore quand il existe plus de patterns (textures et couleurs) dans l'objet (le nombre de points détectés s'augmente).

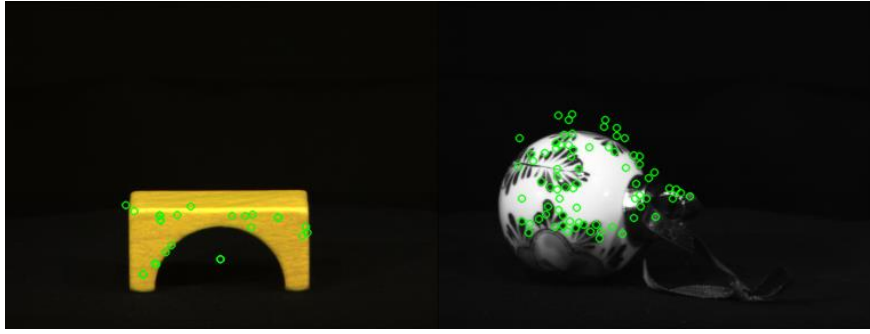


Figure 5 : 251_ c (gauche, 25 keypoints détectés par Brief) et 353_i120 (droite, 75 keypoints détectés par Brief).

6. CONCLUSIONS

- L'ORB a une performance meilleure que d'autres méthodes dans le cas général parce qu'il est capable de détecter plus de points caractéristiques.
- La rotation de l'objet déforme et cache la forme de l'objet. Il est donc difficile à trouver ses images similaires.
- L'ORB est la méthode la plus rapide en termes de temps de calcul parmi les trois.
- Le nombre de points caractéristiques a une influence importante sur la performance, nous constatons qu'il a besoin au moins 100 pour caractériser l'objet dans une image.
- La performance n'est pas très sensible au nombre de clusters (c'est-à-dire, nombre de *visual words* pour calculer les histogrammes). Un nombre de 25, 50 ou 100 n'a pas de différence conséquente sur la performance.
- Les points caractéristiques se situent principalement sur les textures et la forme de l'objet. Un objet ayant une couleur uniforme est plus difficile à extraire des points caractéristiques.