Reconnaissance des formes pour l'analyse et l'interprétation d'images

Rapport TP 1 – 2: SIFT / Bag of words

Etudiant :	**Numéro**
DJEGHRI Amine	3801757
MAMOU Idles	3803676

Octobre 2020

Partie 1 – SIFT

1.1 Calcul du gradient d'une image

1. Les masques de Sobel 3x3 sont séparables en deux vecteurs : hx= [-1,0,1] hy=[1,2,1]
   Le vecteur hx permet de calculer la variation d'intensité (dérivé gradient afin de détecter les contours)
   Le vecteur hy permet d'appliquer un lissage gaussien
2. L'intérêt de séparer le filtre de convolution est d'optimiser les calculs avec moins d'opération, on travaille avec des vecteurs au lieu d'une matrice

1.2 Calcul de la représentation SIFT d'un patch

3. Le rôle de la pondération par masque gaussien est de réduire le bruit pour ne pas avoir de fausses détections (faux contours) causées par ce dernier
4. Calculer le gradient pour 8 directions, et ainsi pouvoir construire l'histogramme, le gradient obtenu après une discrétisation est une approximation du gradient réel, les directions de gradients sont aussi utilisées plus tard pour rendre le descripteur robuste aux rotations
5. - **Norme 2 du descripteur inférieur à un Seuil de 0,5** : éliminer les valeurs à faible contraste ce qui nous permet de dire si des contours sont présents ou pas (Plus le seuil est bas, plus les contours et les caractéristiques non pertinentes de l'image seront détectées, par contre un seuil élevé peut amener à raté des contours)
   - **Seuiller les valeurs supérieures à 0.2** : permettre au descripteur d'être invariant aux changements de luminosité en ramenant toutes les valeurs qui sont supérieures à 0.2 à 0.2.

6. Le principe du SIFT est une façon raisonnable pour faire de l'analyse d'image car : il est robuste aux rotations, luminosité, changement d'échelle et au cadrage, ce qui permet de savoir par exemple si deux images montrent la même chose même si elles ont été prises de deux angles différents
7. Interprétation des résultats :

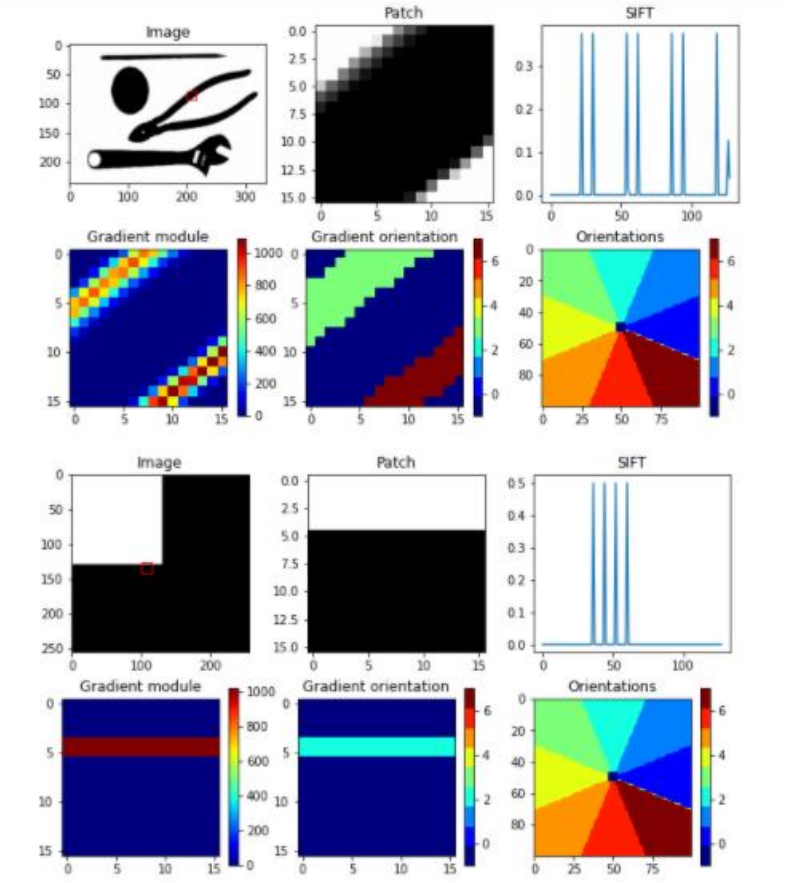

Figure 1.1 : Exemple de représentation

Les résultats nous décrivent les contours, la taille et directions du gradient ainsi que leur localisation

Par exemple pour la deuxième image, nous avons un patch avec un contour horizontal.

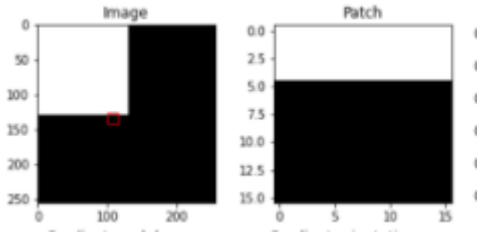

Figure 1.2 : Exemple de l'image 1 et un patch

Le patch est divisé en 16 sous régions (de 1 à 16), et le contour se trouve dans les régions : 5,6,7,8, donc nous avons un gradient dans ces 4 régions et qui a la même direction (le découpage est horizontal pour les 4 sous régions et est le même) et pour les autres régions (de 1 à 4 et de 9 à 16 le gradient est nul (module 0).

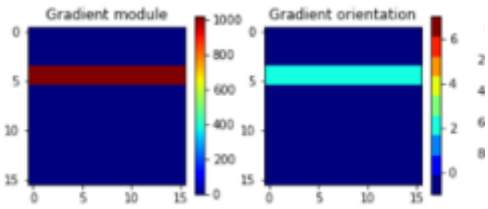

Figure 1.3 : gradient module et orientation

Ceci nous permet de conclure que nous aurons 4 piques.

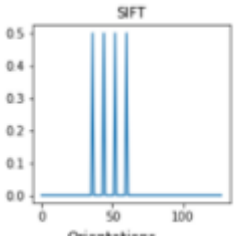

Figure 1.4: Sift de du patch

Le même raisonnement peut être suivi pour interpréter les autres régions (par exemple l'image 1)

Le gradient est non nul pour les sous régions 1,2,5,6 et 12,12,15,16 ce qui nous permet déjà de conclure que nous aurons 8 piques dans la représentation du SIF, sauf que les orientations du gradient des 4 premières seront opposées aux 4 dernières.

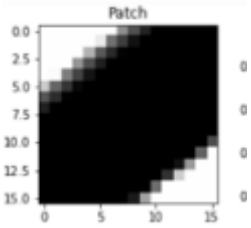

Figure 1.5 : patch tiré à partir de l'image 1

Partie 2 - Dictionnaire visuel

8. Le sac de mot permet de rapprocher et d'identifier les SIFT pris d'une image aux caractéristiques (features) du sac de mot afin de trouver celui qui leurs correspond le mieux.

9. Minimiser => trouver le minimum => gradient =0
   Dérivé de l'équation donnée = 0

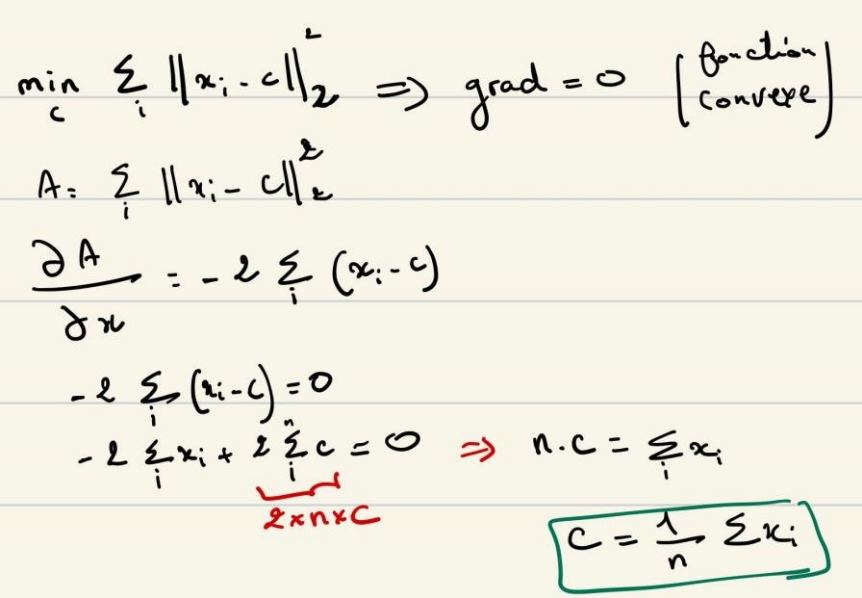

Figure 1.6 : démonstration de la question 9

10. "Elbow" et 'silhouette'
    - Grid search

11. L'analyse des éléments du dictionnaire doit se faire à travers des exemples de patchs et pas directement car :
    Nous ne pouvons pas visualiser les éléments du dictionnaire qui sont des vecteurs qui représentent seulement les caractéristiques, pour visualiser il nous faut des images dont le patch est proche des vecteurs des éléments de notre dictionnaire

12. L'image suivante représente les images proches des 1001 centres de clusters, nous pouvons par exemple voir : une image de lignes verticales, grille d'une fenêtre, mur en brique …ect

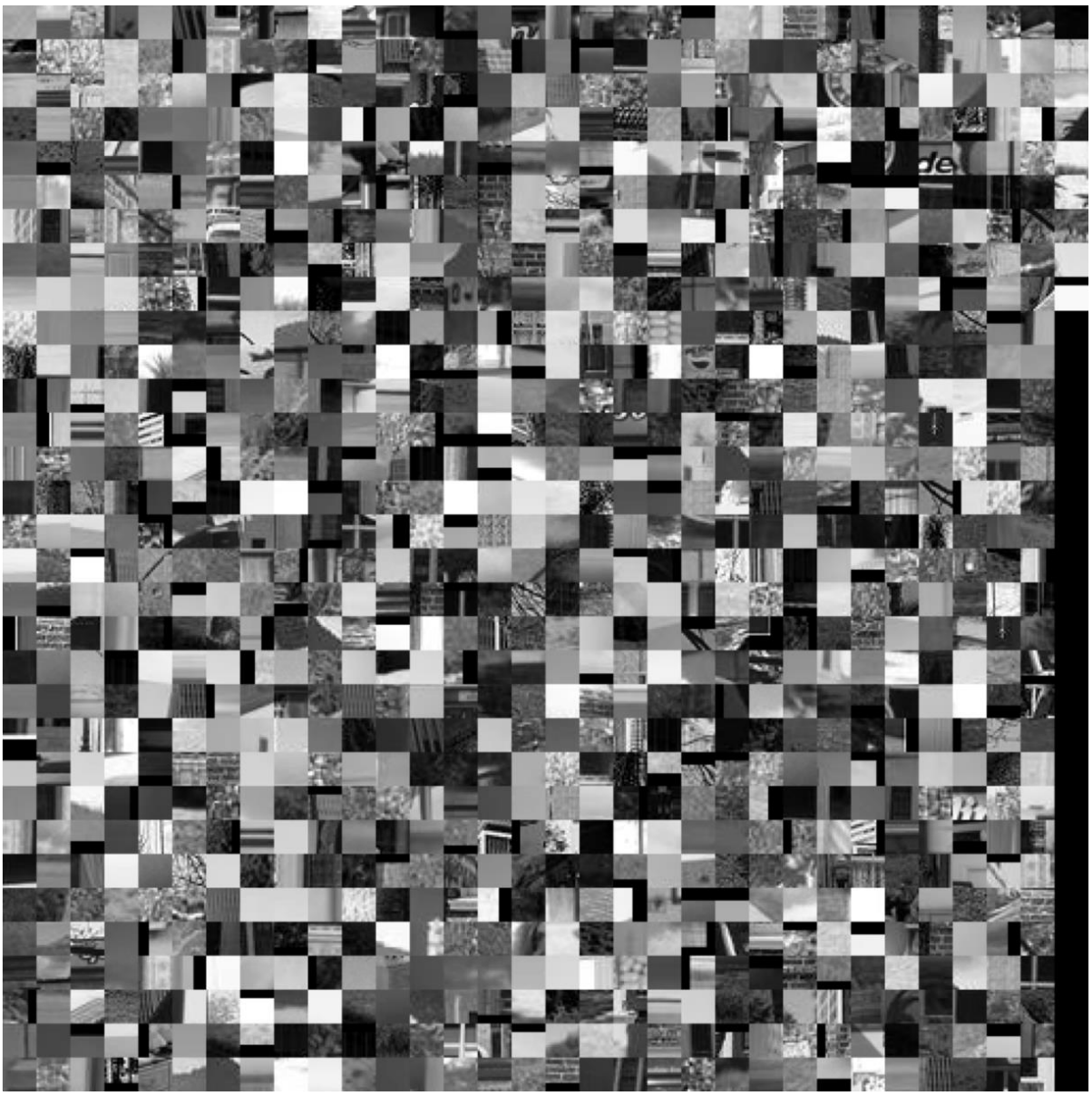

Figure 1.7 : exemples de patchs des centres des 1001 clusters

Partie 3 - Bag of Words (BoW)

13. Z représente la fréquence de chaque cluster détecté dans notre image (l'axe X représente les centres des clusters, l'axe Y représente la fréquence des patchs qui se rapprochent le mieux de ces centres
14. Les résultats obtenus :

Figure 1.8 : Résultats obtenus sur l'image

Nous Pouvons voir que le cluster bleu (en haut à gauche) qui représente une zone uniforme a été associé à la partie 'ciel' de l'image, le cluster vert et violet ont été associés aux feuilles d'arbres, le cluster jaune au cadre des fenetres …ect

15. Interet du codage au plus proche voisin : léger et moins couteux
16. Pooling somme (sum pooling) : favorisent les mots qui apparaissent le plus et est invariant à la position
    Mean pooling : utilise la moyenne (donc divise la somme par le nombre) ceci est presque identique au somme pooling
    Max pooling : prend le mot qui apparait le plus
17. L2 interet : ne varie pas aux changements de rotation donc ne perd pas la direction contrairement à la norme L1