



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

Rapport de TP 1-2

- Réalisé par :

RAMOUL Samy Rayan : *samy.rayan@etu.sorbonne-universite.fr*

30/10/2020

Chapitre 1

SIFT

1.1 Calcul du gradient d'une image

Q1 - Décomposition des masques

Les vecteurs composants des deux matrices sont :

$\begin{bmatrix} 1 & 2 & 1 \\ -1 & 0 & 1 \end{bmatrix}$

Le premier vecteur correspondant à un lissage gaussien. Le second à la dérivée de gradient.

Q2 - Intérêt de la séparation

L'objectif de séparer le filtre de convolution est de réduire le coût de calcul de l'opérateur.

1.2 Calcul de la représentation SIFT d'un patch

Q3 - Role de la pondération par masque gaussien

L'intérêt de la pondération par masque gaussien c'est de permettre de lisser le patch évitant ainsi le bruit qui pourrait fausser la détection de contours par gradient.

Q4 - Role de la discretisation des orientations

Il est important de discrétiser les orientations de gradient afin de préparer l'étape du calcul d'histogramme pour la représentation SIFT, qui serait irréalisable avec des orientations continues.

Q5 - Intérêt des différents post-processings appliqués au SIFT

On calcule la norme du descripteur afin d'obtenir la magnitude du gradient. On le seuille ensuite comme supérieur à 0.5 pour juger pouvoir extraire plus facilement les contours en supprimant les faibles contrastes. Et ensuite on seuille les valeurs supérieures à 0.2 afin de rendre l'algorithme moins sensible aux variations de luminosité.

Q6 - En quoi le principe du SIFT est une façon raisonnable de décrire numériquement un patch d'image pour faire de l'analyse d'image

La technique SIFT se base sur le fait d'utiliser une représentation basée sur les contours et met donc en avant les objets, et ce de manière indépendante à leur orientation ou luminosité ce qui la rend utile dans le domaine de l'analyse de l'image car applicable quelque soit les conditions de prise de l'image.

Q7 - Interprétez les résultats que vous avez obtenus

On observe de par les figures 1.1 et 1.2 que pour ces patches sélectionnés dans des images différentes, les gradients arrivent à efficacement détecter les contours, leurs localisations, tailles et directions.

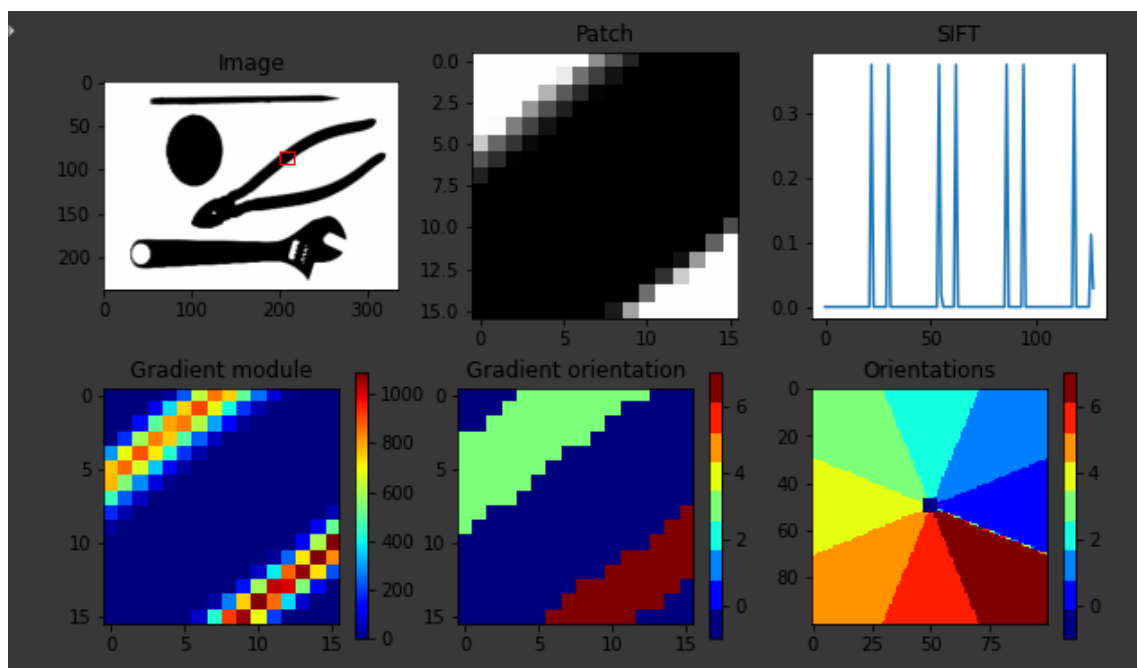


FIGURE 1.1: Premier exemple de SIFT

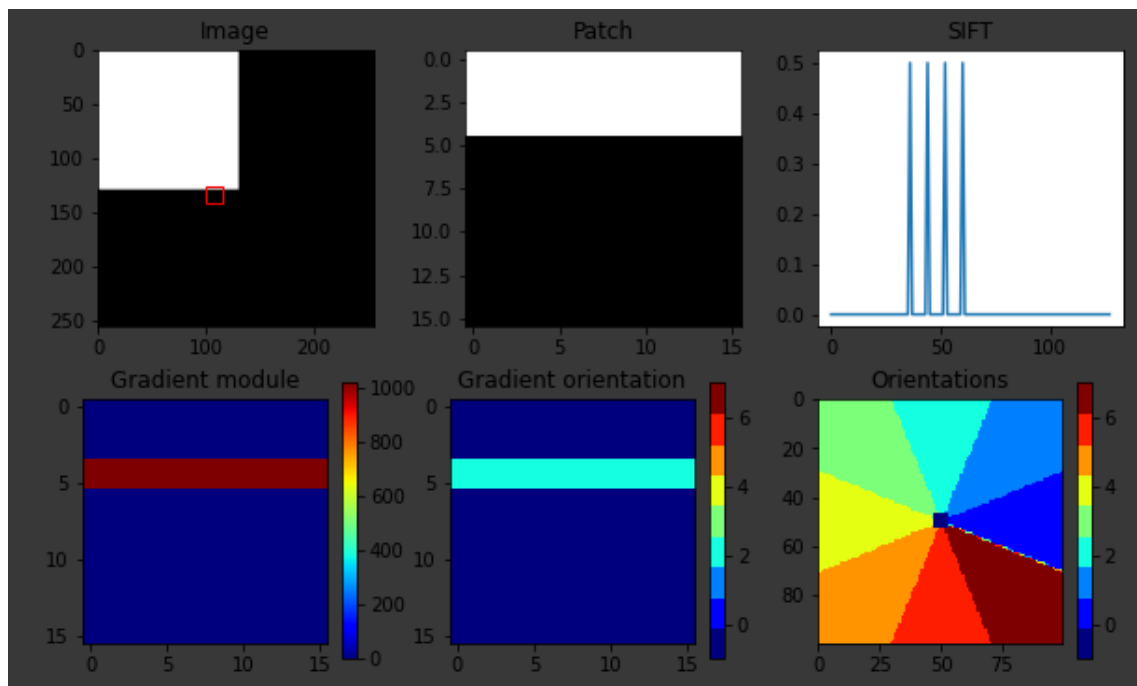


FIGURE 1.2: Second exemple de SIFT

Chapitre 2

Dictionnaire Visuel

Q8 - Nécessité du dictionnaire dans le processus général de reconnaissance d'image que nous sommes entrain de mettre en place

Le dictionnaire est un ensemble de points de référence du Bag of Feature ainsi si le nombre de clusters choisi est 1000, 1000 points de référence de représentation SIFT seront disponibles.

Q9 - Démonstration du centre du cluster

$\{x_i\}_{i=1 \dots n}$ assignés au cluster c

$$f = \min_c \sum_i \|x_i - c\|^2 \Rightarrow \frac{df}{dc} = 0 \text{ (fonction convexe)}$$
$$\frac{\partial \|x_i - c\|^2}{\partial c} = \frac{\partial (x - c)}{\partial c} \cdot \frac{\partial 2(x - c)}{\partial c} = -1 \cdot 2 \cdot (x - c)$$
$$\frac{\partial f}{\partial c} = -2 \sum (x_i - c) = -2(\sum x_i - nc)$$
$$\Rightarrow \sum x_i = nc \Rightarrow c = \frac{\sum x_i}{n}$$

FIGURE 2.1: Demonstration

Q10 - Comment choisir le nombre de clusters idéal

Pour trouver le nombre de clusters idéal on peut utiliser une méthode comme “Silouhette” qui est une mesure de la qualité de représentation (selon le nombre de clusters choisi) et qui est exprimé par la différence entre la distance moyenne avec des points du même groupe et la distance moyenne avec les points des autres groupes.

Q11 - Pourquoi l'analyse des elements du dictionnaire doit se faire à travers des exemples de patchs et pas directement

Les éléments du dictionnaire ne peut se faire qu'au travers d'exemples de patch et pas directement car le dictionnaire est fait d'un ensemble théorique ou parfait de centre de cluster ce qui fait qu'il ne sera pas forcément visuellement parlant.

Q12 - Commentez les résultats que vous aurez obtenus

Après avoir calculer le dictionnaire de référence (centres de clusters SIFT), on prend un echantillon aléatoire d'images et pour plusieurs de leurs regions on calcule les représentations SIFT, et on extrait le sift le plus proche de chaque element du dictionnaire une fois fait on affiche l'image de la region associée.

Cette opération nous permet d'obtenir la figure 2.2, on peut en déduire par conséquent l'association sémantique liée à chaque element du dictionnaire, on remarque notamment une référence pour :

- Une grille de fenêtre.
- Sommet d'arbre.
- Bordure d'escalier.
- Des transitions de zones claires vers des zones sombres (et ce dans plusieurs directions différentes).
- ...ect.

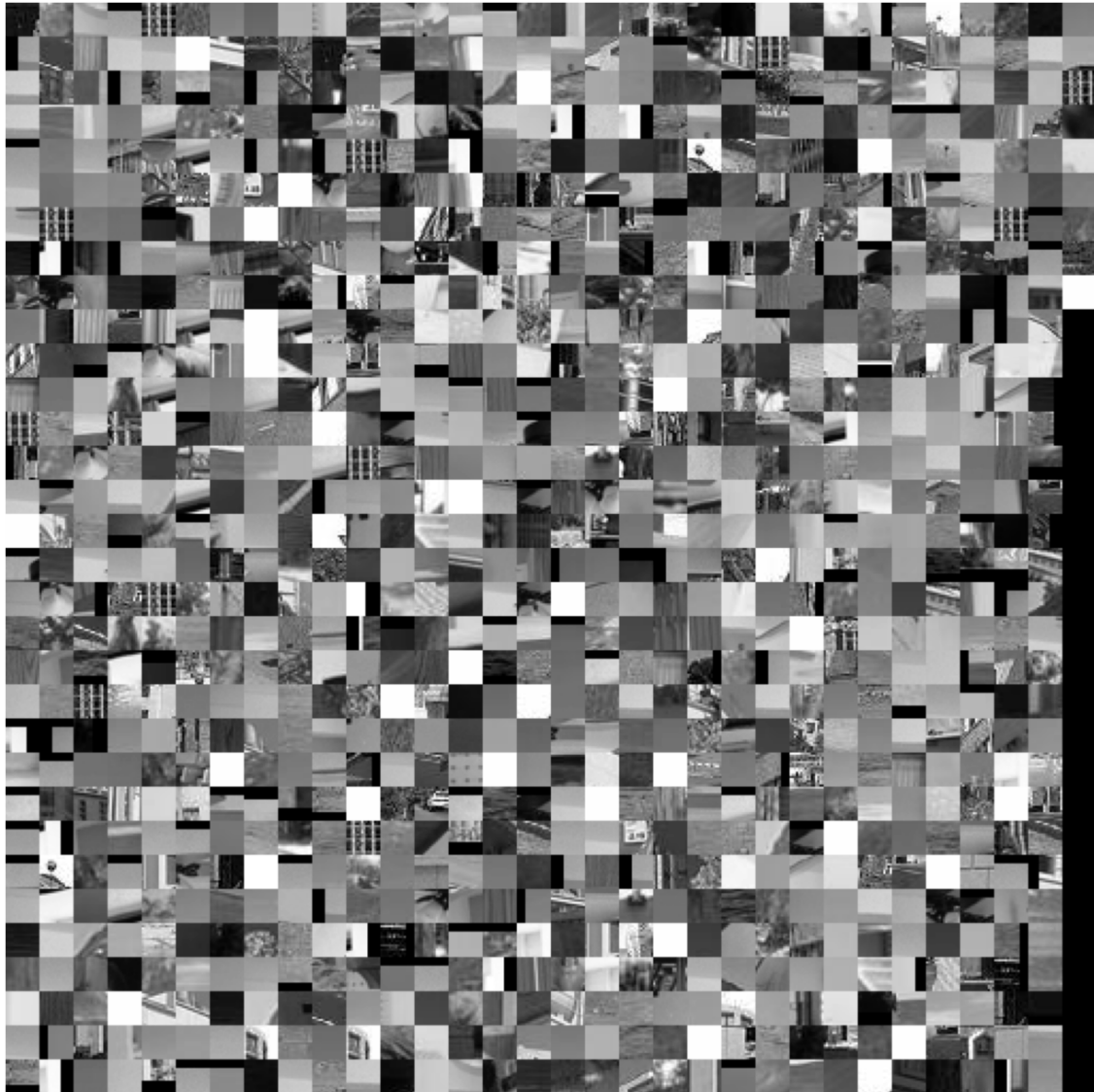


FIGURE 2.2: Regions proches

Chapitre 3

Bag of Words

Q13 - Que représente concrètement notre vecteur pour une image

Pour une image I , le vecteur z représente un histogramme ou l'axe des X représente l'ensemble des centres théoriques des clusters du dictionnaire. Et sur l'axe $Y=f(x)$ le nombre de patches de l'image qui sont les plus proches du centre de cluster x .

Q14 - Montrez et discutez les résultats visuels obtenus

On observe dans la figure 3.1 que pour l'image donnée toutes les régions du ciel ont été détectées comme plus proches d'un seul et même SIFT de référence (couleur orange), c'est le même cas pour les feuillages de buissons ou d'arbres qui eux se retrouvent associés potentiellement à deux références différentes (couleur bleu ou vert) ou encore que les parties de fenêtres sont elles associées à un même SIFT (couleur rose).

Si on prend tout autre image comme par exemple dans la figure 3.3 on observe ces mêmes associations de types de régions aux mêmes références, on peut donc en conclure que cet algorithme parvient à une certaine cohérence, stabilité d'une image à une autre, et à déterminer des classes véritablement différentes d'éléments récurrents dans les images.

Enfin, sur l'histogramme associé (figure), ces mêmes zones (ciel, feuillages ..ect) très représentés sont les plus incrémentés/élevés.



FIGURE 3.1: Regions proches



FIGURE 3.2: Regions proches

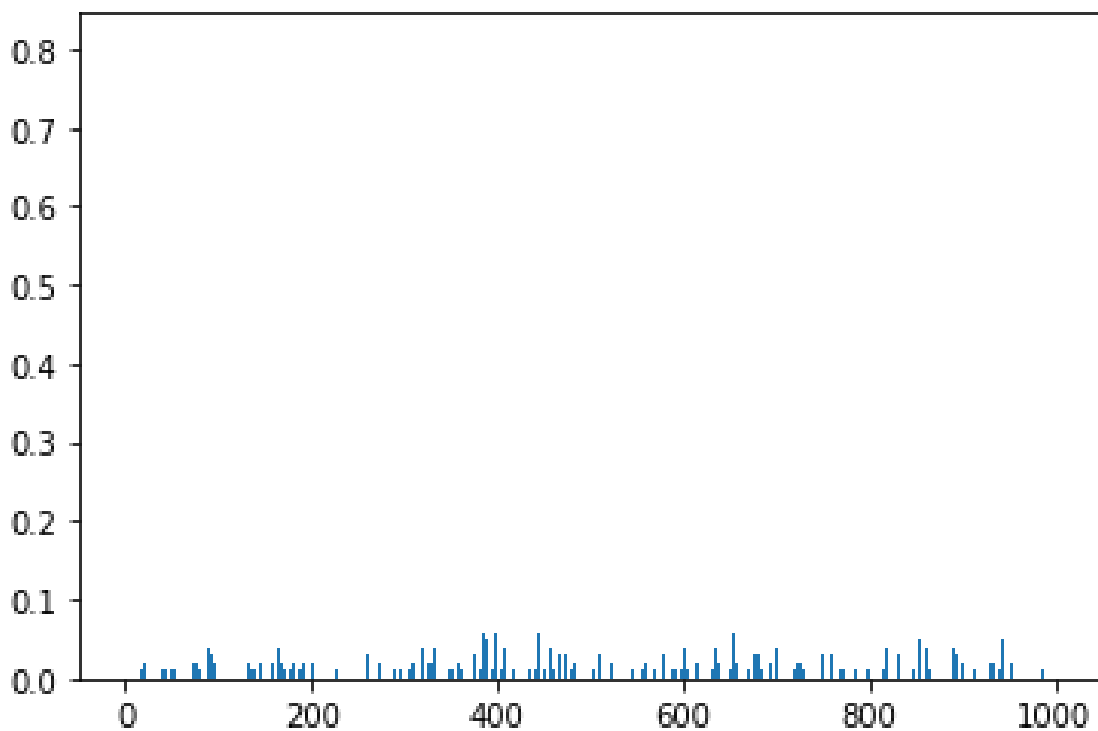


FIGURE 3.3: Histogramme des régions proches de chaque référence SIFT

Q15 - Quelle est l'intérêt du codage au plus proche voisin ? Quel(s) autre codage pourrait-on utiliser ?

L'intérêt du codage au plus proche est d'avoir une représentation assez légère et définissant. On pourrait aussi utiliser la distance au plus proche voisin.

Q16 - Quelle est l'intérêt du pooling somme ? Quel(s) autre pooling pourrait-on utiliser ?

Si on opte pour une représentation par distance au plus proche voisin (au lieu du one-hot-encoding) on pourrait alors opter pour un pooling par moyenne.

Q17 - Quelle est l'intérêt de la normalisation L2 ? Pourrait-on utiliser une autre normalisation ?

La norme L2 est invariante à la rotation et donc ne perd pas la direction tandis que L1 la perd.