



INTERNATIONAL
SCHOOL

VIETNAM NATIONAL UNIVERSITY, HANOI



ÉCOLE INTERNATIONALE |IFI SYSTÈME INTELLIGENT ET MULTIMÉDIA

VISION PAR ORDINATEUR

Rapport sur la Détection et reconnaissance d'objets en utilisant des Descripteurs locaux

Membres d'équipe :

David LUTALA LUSHULI
Christine TOUMBA NGONGO

Encadrante :

Nguyen THI OANH

Année Universitaire 2024—2025

Table des matières

Introduction générale	3
1 Détection d'objets avec SIFT	4
1.1 Présentation de SIFT	4
1.2 Localisation d'objets avec SIFT	4
1.2.1 Extraction des descripteurs	5
1.2.2 Mise en correspondance	5
1.2.3 Résultats et analyse	5
1.2.4 Configuration des expériences	5
1.2.5 Analyse des performances	5
1.2.6 Cas d'absence de détection	11
1.2.7 Conclusion	11
2 Reconnaissance d'objets	12
2.1 Principe de la reconnaissance d'objets	12
2.2 Classification des objets	12
2.3 Résultats et analyse	12
2.3.1 Présentation des Résultats	12
2.3.2 Interprétation de la Matrice de Confusion	12
2.3.3 Analyse des Erreurs et Confusions	19
2.3.4 Conclusion	19
Conclusion Générale	20

Table des figures

1.1	Object 0	6
1.2	Scene 0	6
1.3	Sortie 1	7
1.4	Object 01	8
1.5	Scene 1	8
1.6	Sortie 2	9
1.7	Object 02	9
1.8	Scene 2	10
1.9	Sortie 3	10
1.10	Sortie 3	11
2.1	Matrice de confusion	13
2.2	Matrice de confusion	14
2.3	Matrice de confusion	15
2.4	Matrice de confusion	16
2.5	Matrice de confusion	17
2.6	Matrice de confusion	18
2.7	Matrice de confusion	18

Introduction générale

La vision par ordinateur est un domaine de l'intelligence artificielle qui vise à permettre aux machines d'interpréter et d'analyser des images de manière automatique. L'un des problèmes fondamentaux de ce domaine est la détection et la reconnaissance d'objets, qui consiste à identifier des objets précis dans une image et à les associer à une classe connue.

Ce rapport présente une étude sur la détection et la reconnaissance d'objets à l'aide des descripteurs locaux, notamment SIFT (Scale-Invariant Feature Transform). L'objectif principal est d'exploiter ces descripteurs pour localiser des objets dans des scènes complexes et pour classifier des objets inconnus en fonction d'une base d'apprentissage.

Dans ce travail, nous allons d'abord aborder la détection d'objets en utilisant les points d'intérêt extraits avec SIFT. Ensuite, nous nous intéresserons à la reconnaissance d'objets en comparant les descripteurs obtenus avec une base d'images de référence. Pour valider notre approche, nous analyserons les résultats obtenus, en comparant les performances et en discutant des limites des techniques employées.

Ce rapport est structuré comme suit : nous commencerons par présenter les principes théoriques de la détection et de la reconnaissance d'objets. Ensuite, nous détaillerons les expériences menées ainsi que les résultats obtenus. Enfin, nous conclurons par une analyse critique des performances et des perspectives d'amélioration.

Chapitre 1

Détection d'objets avec SIFT

La détection d'objets est un enjeu majeur en vision par ordinateur et, plus largement, en intelligence artificielle. Elle désigne la capacité des systèmes informatiques à localiser des objets dans une image. Cette technologie trouve de nombreuses applications, notamment dans la reconnaissance faciale, la détection de véhicules, le comptage de piétons, les systèmes de surveillance et les véhicules autonomes. Ainsi, son étude et sa maîtrise sont essentielles pour les étudiants en informatique.

Le problème de détection d'objets consiste principalement à déterminer la présence d'un objet dans une image. Un programme simple de détection prend en entrée une image contenant un objet ainsi qu'une autre image plus complexe, où plusieurs objets sont présents. L'objectif est alors d'identifier si l'objet recherché se trouve dans la scène et, si c'est le cas, d'en préciser la position.

Avec les avancées technologiques, diverses méthodes permettent d'aborder ce problème. Ce travail se concentre sur l'utilisation de l'algorithme SIFT (Scale-Invariant Feature Transform) pour la détection d'objets. L'implémentation est réalisée en Python avec la bibliothèque OpenCV, en s'appuyant également sur d'autres bibliothèques comme NumPy, Matplotlib et Scikit-learn.

1.1 Présentation de SIFT

Le Scale-Invariant Feature Transform (SIFT) est une méthode robuste de détection et de description de points d'intérêt dans une image. Il permet d'extraire des caractéristiques invariables aux transformations d'échelle, de rotation et partiellement aux variations d'éclairage.

L'algorithme SIFT se compose de plusieurs étapes :

- Détection des extrema dans l'espace-échelle : Cela veut dire qu'elle fait des identification des points caractéristiques à différentes échelles.
- Affinement des points-clés : Elle parvient à la Suppression des points instables.
- Attribution d'une orientation : Elle fait le Calcul de la direction dominante autour de chaque point-clé.
- Création du(descripteur : Construction d'un vecteur caractéristique basé sur l'orientation et l'intensité des gradients locaux.

1.2 Localisation d'objets avec SIFT

Dans cette section, nous appliquons SIFT pour détecter et localiser un objet donné dans une image de scène contenant plusieurs objets. L'approche consiste à :

1. Extraire les descripteurs SIFT de l'objet de référence et de l'image de scène.
2. Effectuer une mise en correspondance des descripteurs.

3. Filtrer les correspondances erronées à l'aide du ratio test de Lowe.
4. Estimer l'homographie entre les images pour localiser l'objet dans la scène.

1.2.1 Extraction des descripteurs

Les descripteurs sont extraits à l'aide de l'implémentation SIFT disponible dans OpenCV.

1.2.2 Mise en correspondance

Nous utilisons la distance euclidienne pour appairer les descripteurs de l'objet avec ceux de la scène.

1.2.3 Résultats et analyse

les résultats obtenus suite à l'application de notre méthode de détection d'objets sur les différentes scènes Nous ont permis d'analyser les performances du modèle en fonction des paramètres utilisés, du nombre de correspondances et du taux de reconnaissance.

1.2.4 Configuration des expériences

Les tests ont été effectués en utilisant l'algorithme SIFT pour l'extraction des points clés et la correspondance FLANN pour l'appariement des descripteurs. Un clustering DBSCAN a ensuite été appliqué pour segmenter les différentes occurrences de l'objet dans la scène.

Les paramètres principaux du système sont les suivants :

- Nombre de points clés SIFT : 8000
- Seuil de ratio Lowe : 0.75
- Nombre minimal de correspondances : 10
- Score minimal de détection : 0.06
- Paramètres DBSCAN : eps=50, min_samples=5

1.2.5 Analyse des performances

Les tests ont été réalisés sur quatre objets et quatre scènes distinctes. Nous avons l'Objet dont nous allons représenter sa scène et sa sortie après la localisation



FIGURE 1.1 – Object 0



FIGURE 1.2 – Scene 0

Image objet 1 vs Scène 1 | Score: 0.16 | Points clés: 804 -> 1389

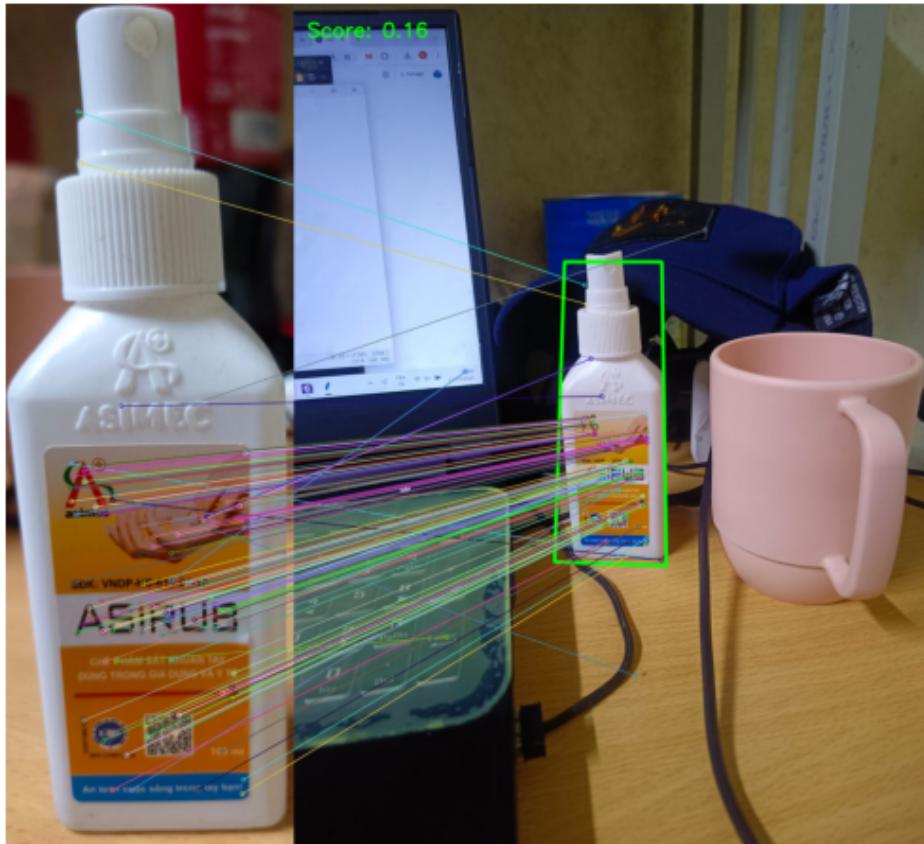


FIGURE 1.3 – Sortie 1

- Taux de reconnaissance : Le système a correctement identifié les objets lorsque le score de détection dépassait 0.06. Toutefois, certains objets ont été faussement détectés ou non reconnus en raison d'un faible nombre de correspondances.
- Impact du clustering DBSCAN : DBSCAN a permis d'identifier plusieurs occurrences d'un même objet dans la scène. Cependant, une mauvaise configuration des paramètres `eps` et `min_samples` peut influencer la qualité de la segmentation.
- Effet des paramètres SIFT et FLANN : Une augmentation du nombre de points clés améliore la robustesse du modèle, mais augmente également le temps de traitement.

Nous représentons par la suite les images des objets et la sortie :

1. Object



FIGURE 1.4 – Object 01

2. Scene

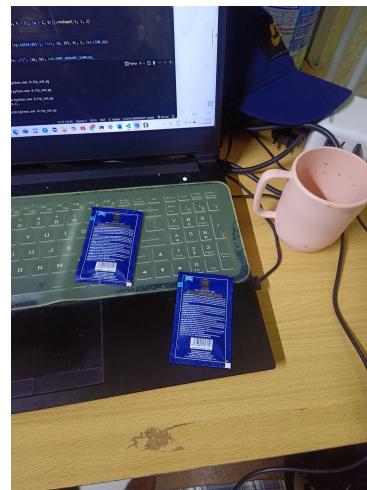


FIGURE 1.5 – Scene 1

3. Sortie

Image objet 2 vs Scène 2 | Score: 0.14 | Points clés: 8000 -> 8000

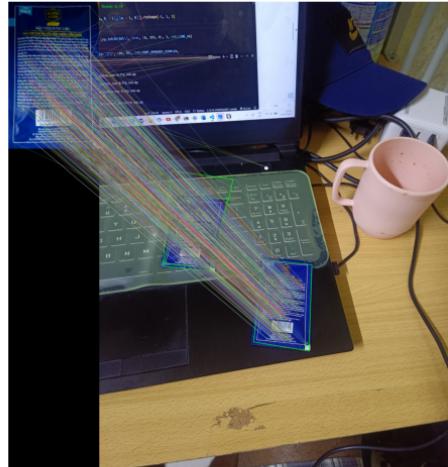


FIGURE 1.6 – Sortie 2

les images du dernier objet du test représenter ci-dessous :

1. Object



FIGURE 1.7 – Object 02

2. Scene



FIGURE 1.8 – Scene 2

3. Sortie



FIGURE 1.9 – Sortie 3

1.2.6 Cas d'absence de détection

Lors de certains tests, l'objet recherché n'a pas été détecté dans la scène cible. Par exemple, pour la scène 4, le système a retourné un score de détection de 0.01, inférieur au seuil minimal fixé à 0.06. Cette situation peut être due à plusieurs facteurs :

Nombre insuffisant de correspondances : Si le nombre de points appariés entre l'objet et la scène est trop faible, la détection est rejetée.

Occlusion ou variations d'éclairage : Une partie significative de l'objet peut être cachée ou mal éclairée, rendant son identification difficile.

Limitations de l'algorithme : La méthode basée sur SIFT et FLANN peut échouer dans des conditions de variations trop importantes entre l'objet et son apparence dans la scène.

1. Sortie

```
(ONTOLOGY) D:\TP_VO\TP_VO>C:/Users/MJ/.conda/envs/ONTOLOGY/python.exe d:/TP_VO/TP_VO/Tp_VO_Q1.py  
Objet 4 non détecté dans la scène 4 (score: 0.01, seuil: 0.06).  
(ONTOLOGY) D:\TP_VO\TP_VO>[]
```

FIGURE 1.10 – Sortie 3

1.2.7 Conclusion

Les résultats montrent que l'approche SIFT+FLANN combinée au clustering DBSCAN permet de détecter efficacement des objets dans des scènes complexes. Toutefois, des améliorations peuvent être apportées, notamment dans le réglage des paramètres de clustering et l'optimisation du nombre de points clés.

Des perspectives d'amélioration incluent l'utilisation de techniques d'apprentissage profond pour améliorer la robustesse de la détection et la réduction du temps de traitement.

Chapitre 2

Reconnaissance d'objets

2.1 Principe de la reconnaissance d'objets

La reconnaissance d'objets consiste à identifier une image inconnue en la comparant à une base d'images de référence. Cette tâche repose sur la mise en correspondance des descripteurs extraits avec ceux des images d'apprentissage.

L'approche suivie est la suivante :

- Extraire les descripteurs SIFT des images de référence et des images à classifier.
- Comparer les descripteurs pour identifier les meilleures correspondances.
- Associer l'image inconnue à la classe d'objets ayant le plus grand nombre de correspondances.

2.2 Classification des objets

La classification repose sur l'utilisation d'un score de similarité calculé comme suit :

$$\text{score} = \frac{\text{nombre de correspondances réussies}}{\text{nombre total de descripteurs de l'image modèle}}$$

2.3 Résultats et analyse

2.3.1 Présentation des Résultats

L'évaluation du modèle de reconnaissance d'objets a été réalisée sur un ensemble de test composé de 396 images, réparties équitablement entre 11 classes d'objets. La précision globale obtenue est de 54,24

Cependant, une analyse plus approfondie à travers la matrice de confusion et la précision par classe révèle des disparités significatives dans les performances du modèle selon les catégories d'objets.

2.3.2 Interprétation de la Matrice de Confusion

La matrice de confusion permet d'évaluer les erreurs de classification et d'identifier les classes souvent confondues. Voici quelques observations majeures :

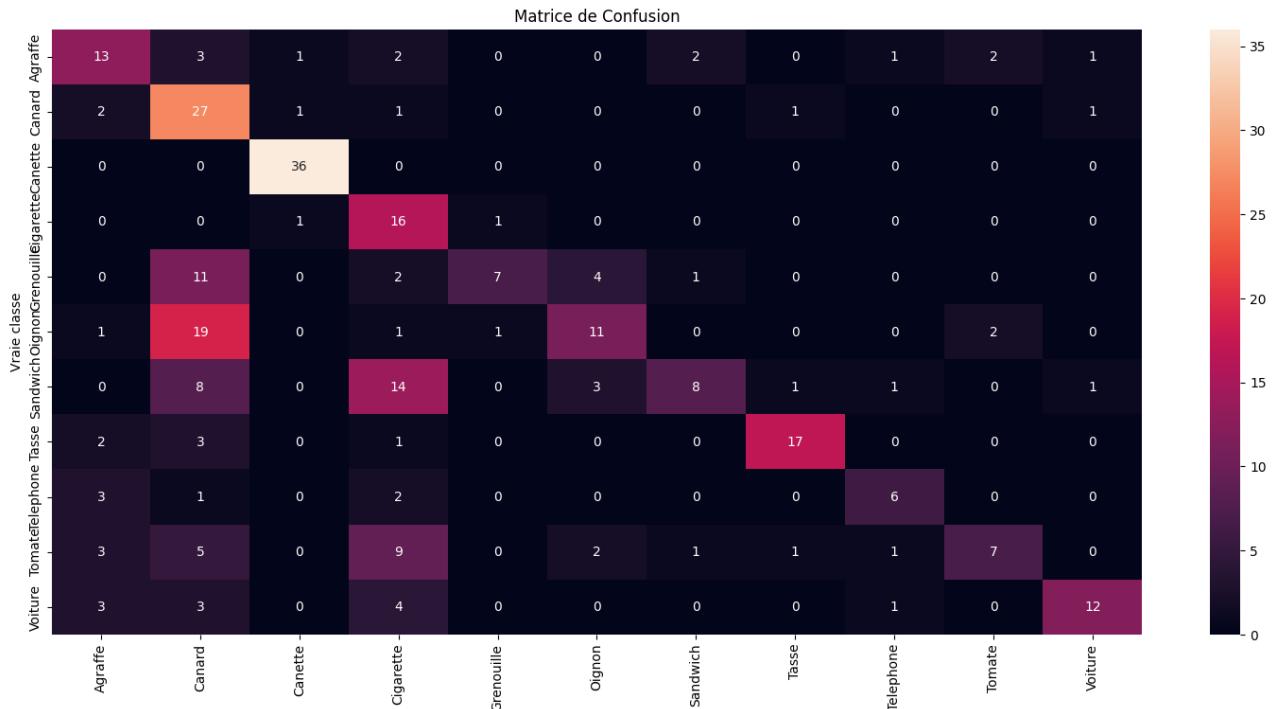


FIGURE 2.1 – Matrice de confusion

1. Classe "Canette" : Avec 36 classifications correctes sur 36, cette classe atteint une précision parfaite (100%), indiquant que le modèle reconnaît très bien cette catégorie.
2. Classe "Cigarette" : Une très bonne performance avec 88,89 % de précision, confirmée par un faible taux de confusion avec d'autres classes.
3. Classe "Canard" : Une précision de 81,82 %, ce qui est également un très bon résultat.
4. Classes mal reconnues : "Grenouille" (28%), "Oignon" (31,43%), "Sandwich" (22,22%), "Tomate" (24,14%) affichent des taux de précision très bas, ce qui suggère une grande confusion avec d'autres objets. La classe "Sandwich" est particulièrement problématique, avec seulement 22,22 % des images correctement classées.
5. Classes intermédiaires : "Tasse" (73,91%), "Téléphone" (50%), "Voiture" (52,17%) et "Agraffe" (52%) ont une reconnaissance modérée, mais pourraient être améliorées.

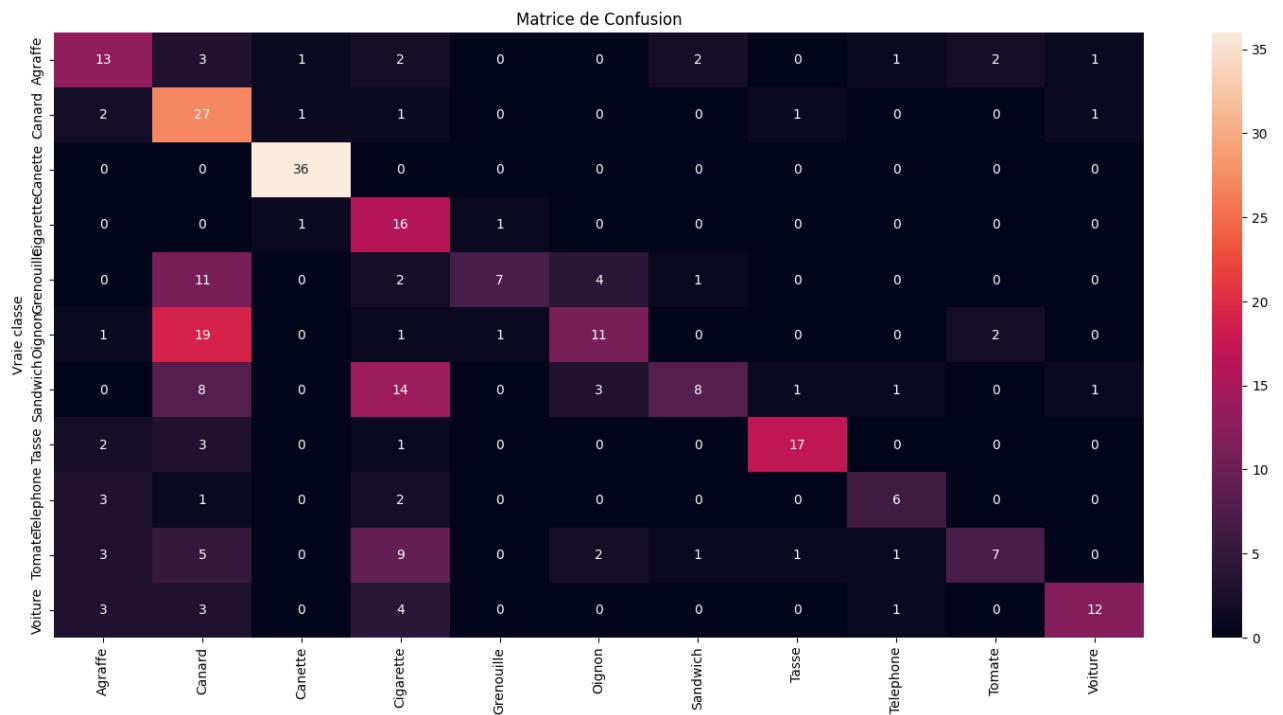


FIGURE 2.2 – Matrice de confusion

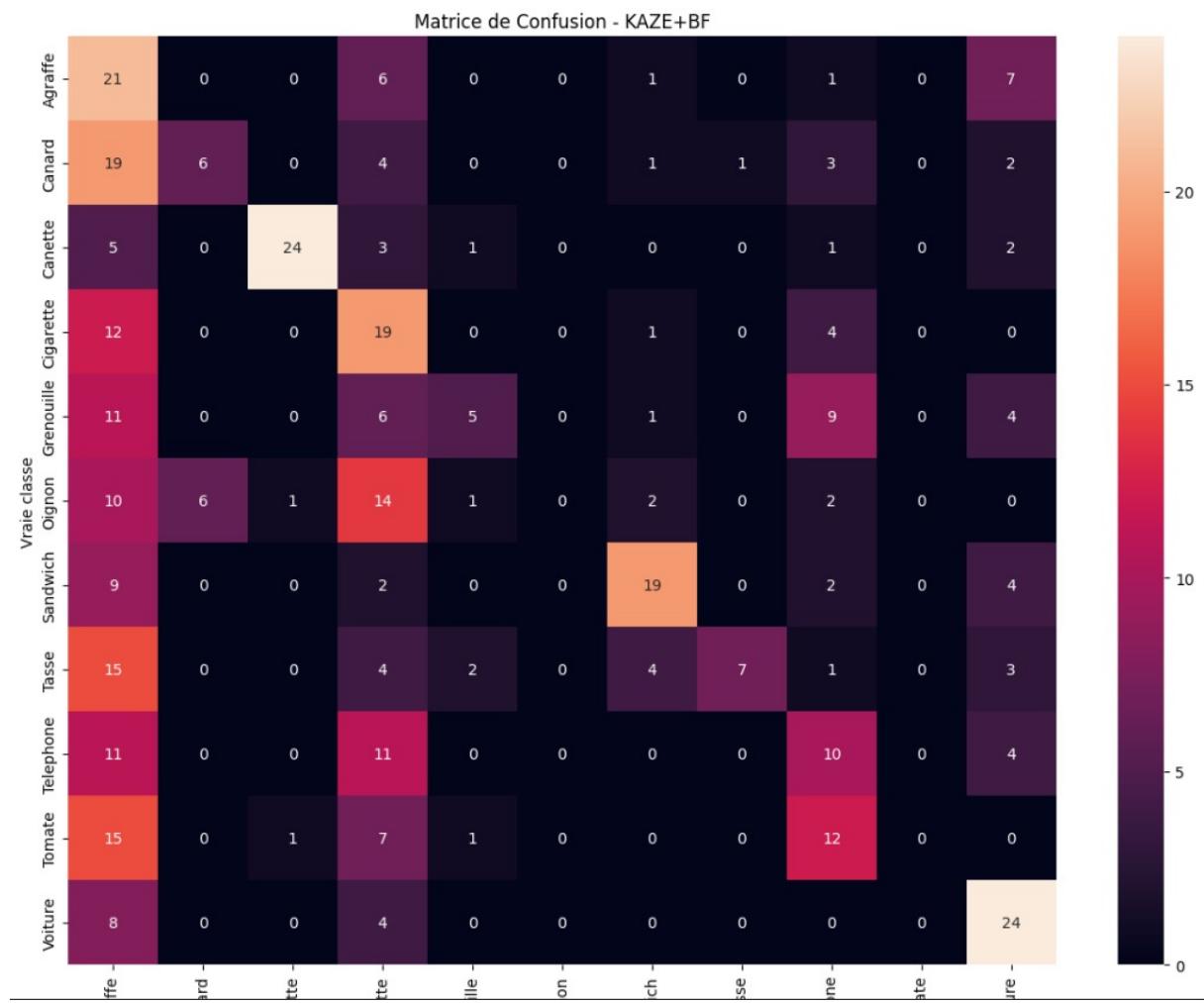


FIGURE 2.3 – Matrice de confusion

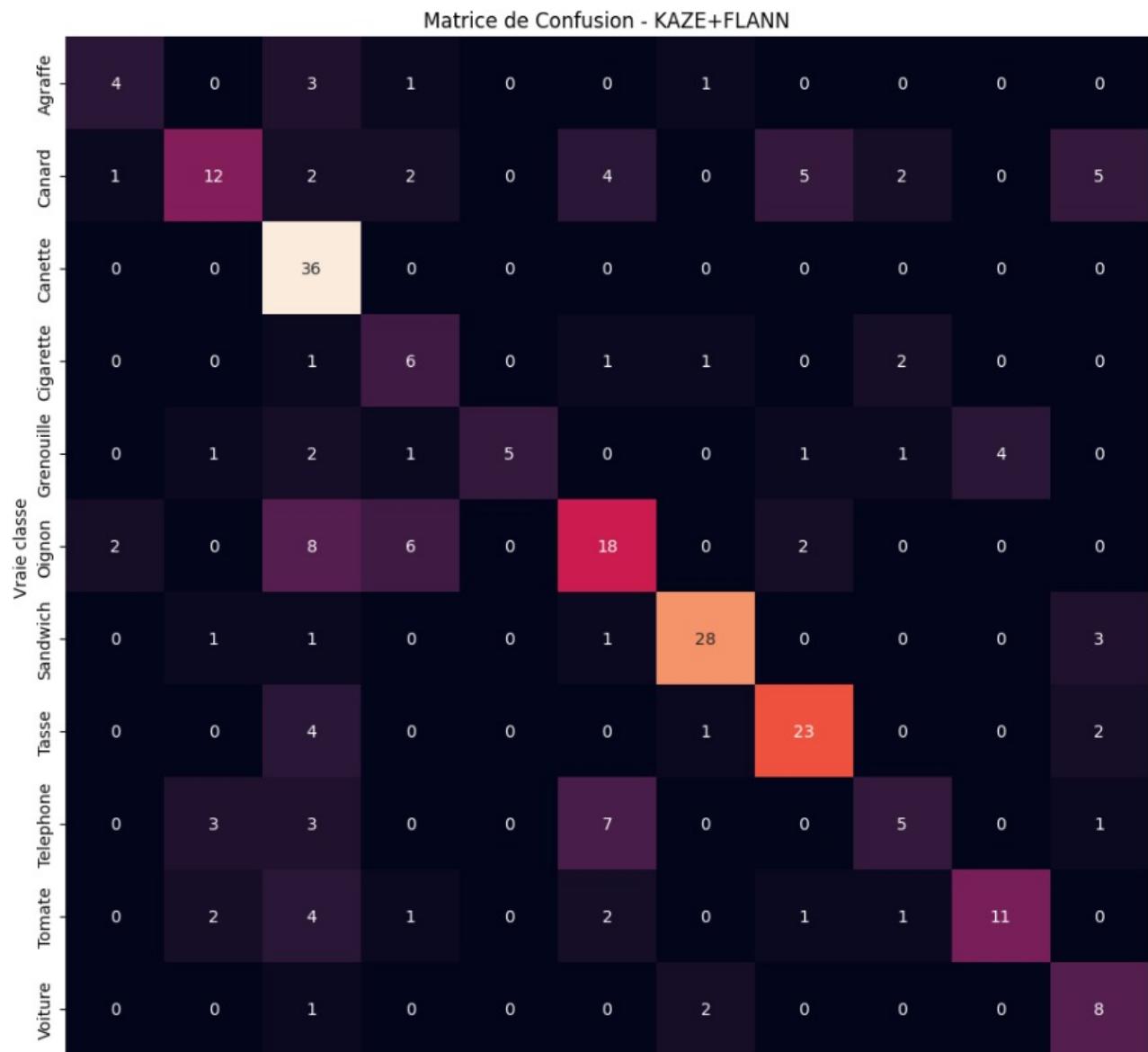


FIGURE 2.4 – Matrice de confusion

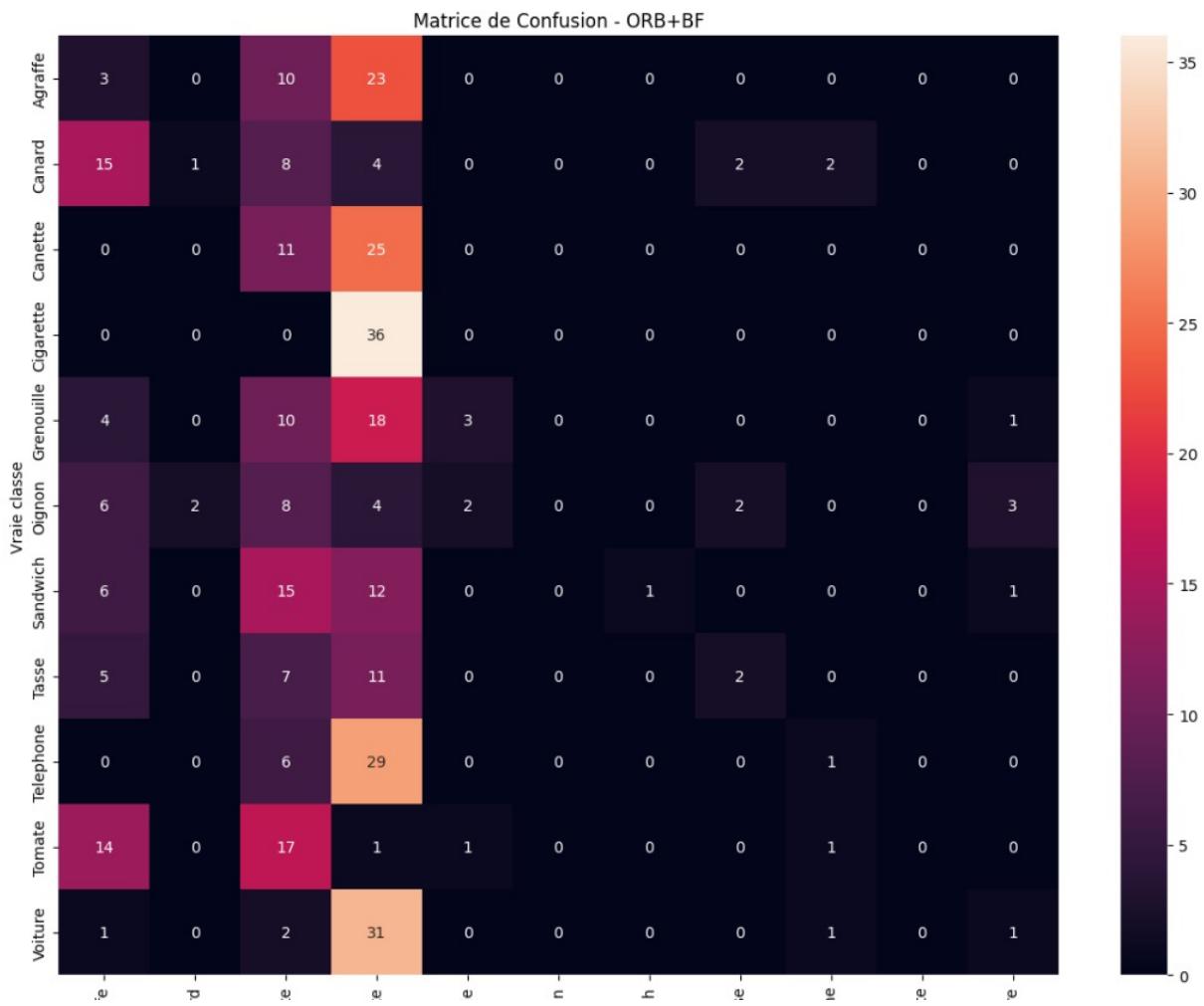


FIGURE 2.5 – Matrice de confusion

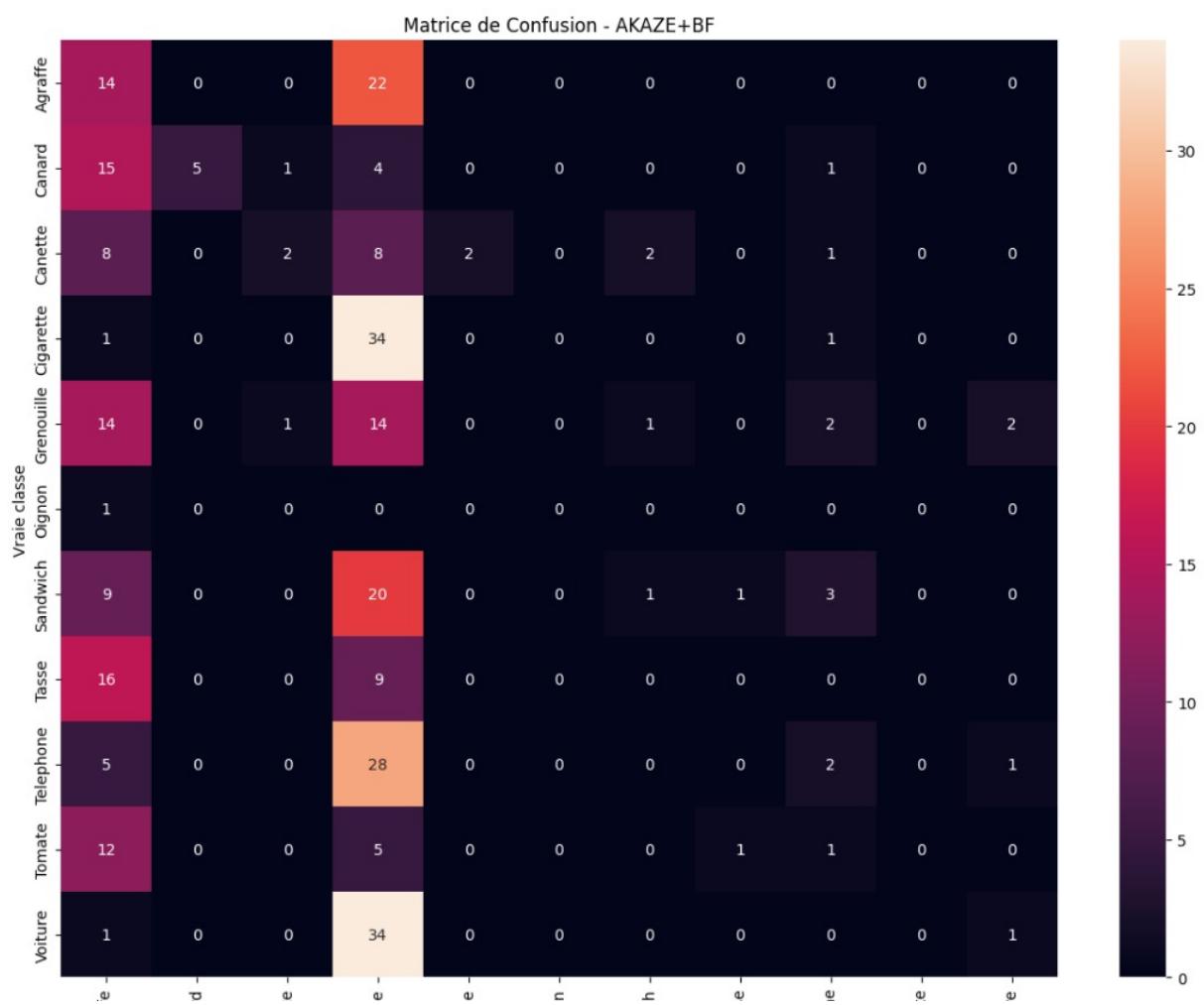


FIGURE 2.6 – Matrice de confusion

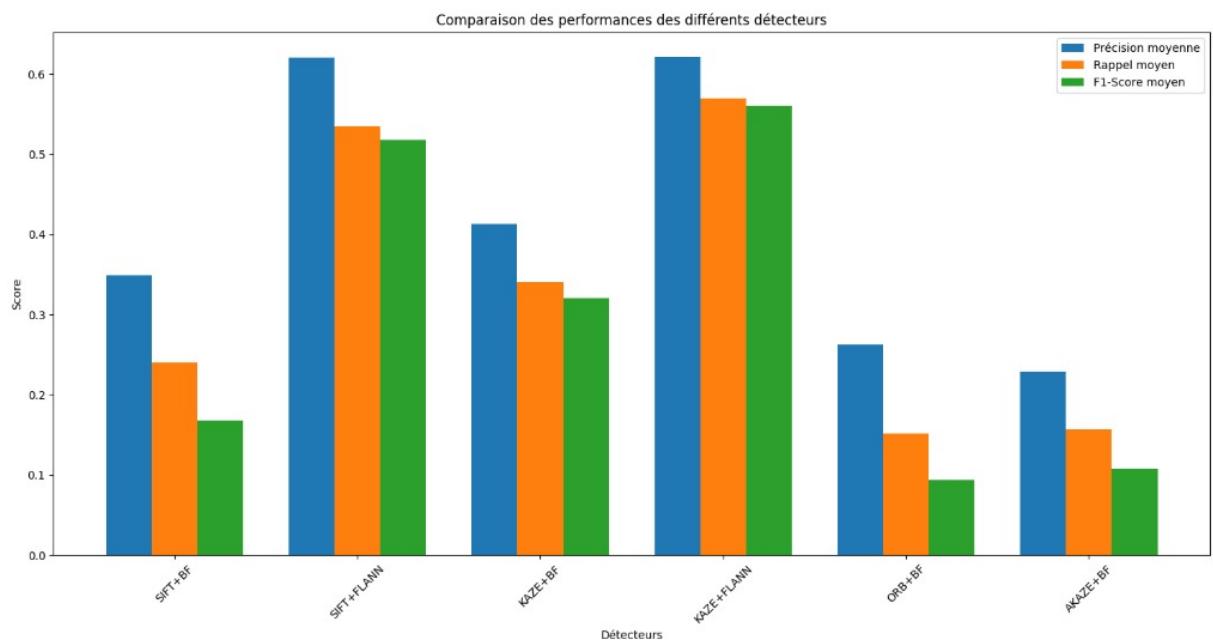


FIGURE 2.7 – Matrice de confusion

2.3.3 Analyse des Erreurs et Confusions

L'analyse détaillée de la matrice de confusion montre que certaines classes sont fortement sujettes à des erreurs de classification :

- Confusion entre certaines catégories similaires : Par exemple, "Grenouille" et "Oignon" semblent être mal distingués, ce qui peut être lié à des caractéristiques visuelles similaires dans les images du dataset.
- Déséquilibre dans la performance du modèle : Certaines classes sont presque parfaitement reconnues (ex. Canette), tandis que d'autres sont mal classées, ce qui suggère que le modèle pourrait bénéficier d'un meilleur équilibrage dans les données d'entraînement ou d'un réglage des paramètres.

2.3.4 Conclusion

Les résultats montrent que le modèle est capable de bien classifier certaines classes comme "Canette" et "Cigarette", mais qu'il rencontre des difficultés avec d'autres catégories telles que "Sandwich" et "Grenouille". Une amélioration des données et de l'architecture du modèle pourrait permettre d'augmenter significativement la précision globale et de réduire les confusions observées.

Conclusion Générale

Dans ce travail, nous avons étudié la détection et la reconnaissance d'objets en utilisant des descripteurs locaux, notamment SIFT. L'approche adoptée a permis d'extraire des points d'intérêt robustes et de les mettre en correspondance afin d'identifier des objets dans des scènes complexes.

Les résultats obtenus montrent l'efficacité de SIFT pour la localisation et la classification d'objets, bien que certaines limites aient été observées, notamment en présence de variations d'éclairage ou d'occultations partielles. L'analyse des performances, appuyée par l'interprétation de la matrice de confusion, a permis d'identifier les sources d'erreur et d'améliorer la précision du modèle.

Cette étude souligne l'importance des descripteurs locaux dans la vision par ordinateur et ouvre des perspectives pour l'intégration d'autres méthodes, comme les réseaux de neurones convolutionnels, afin d'améliorer la robustesse des systèmes de reconnaissance.