Projet 2 - Classification des Scènes Naturelles

AUTEURS DU RAPPORT: **Ginel Dorleon, Gervais Sikadie F., Ednalson Eliodor G.(p21)**
MODULE: **Reconnaissance des Formes**
OPTION: **Systèmes Intelligents et Multimédia**
SUPERVISEUR: **HO Tuong Vinh**
Institut Francophone International

Mot clés: Scènes naturelles, classification, descripteur, HOG, SIFT, classification, K-means, SVM

I. INTRODUCTION

La reconnaissance de scènes naturelles est l'une des tâches marquantes de la vision par ordinateur, permettant de définir un contexte de reconnaissance d'objet. Une grande partie des progrès récents dans la vision par ordinateur a été faite en concevant des fonctions d'image robustes pour les tâches de reconnaissance telles que la reconnaissance d'objets et de scène naturelle. Presque toutes ces fonctionnalités sont basées sur une sorte de propriétés d'image de bas niveau. Avec l'aide des modèles statistiques de plus en plus sophistiqués, ces caractéristiques ont atteint un bon succès dans les tâches de reconnaissance de haut niveau. Particulièrement, le récent développement de détecteurs robustes à un seul objet en fonction de ces caractéristiques est un grand pas. En utilisant ces détecteurs, les chercheurs ont développé des algorithmes pour intégrer davantage des informations de contexte telles que la disposition de scène, l'arrière-plan, la classe et la cooccurrence d'objets pour obtenir de meilleures détections d'objets dans les scènes. C'est ainsi que la classification des scènes naturelles est un problème intéressant dans la vision par ordinateur. Dans ce rapport, nous considérons le problème de la reconnaissance de la catégorie sémantique d'une image. Par exemple, nous pouvons vouloir classer une photo comme représentant une scène (forêt, rue, bureau, etc.) ou contenant un certain objet. Ainsi, on va se proposer de trouver la classification sémantique, par exemple, l'intérieur contre l'extérieur, artificielle contre nature, plage contre désert, d'images arbitraires a été largement étudiée au cours de la dernière décennie.
Dans ce rapport, nous présentons une approche combinant les descripteurs globaux et locaux pour faire la classification, nous totalisons un tauxde 81% sur la base 13 Natural Scene Categories [6].

II. OBJECTIF

Dans cette étude, notre objectif se porte sur l'expérimentation et la classification des scènes naturelles. Ainsi dans ce rapport, nous présentons un état de l'art, l'implémentation de l'approche de notre solution proposée, les résultats obtenus ainsi qu'une analyse détaillée critique et comparative des résultats.

III. CONTEXTE

La capacité à analyser et à classer avec précision et rapidité la scène dans laquelle nous nous trouvons est très utile dans la vie courante. Des chercheurs ont constaté que les humains ont capables de classer les scènes naturelles complexes contenant des animaux ou des véhicules très rapidement [2]. Li et ses collègues ont montré que peu ou pas d'attention est nécessaire pour la catégorisation de la scène naturelle rapide. Mais cependant, pour comprendre le contexte d'une scène complexe, il faut d'abord reconnaître les objets et ensuite reconnaître la catégorie de la scène. Peut-on reconnaître le contexte d'une scène et la classer sans avoir reconnu d'abord les objets présents ?

Un nombre des études récentes ont présenté des approches pour classer l'intérieur contre l'extérieur, la ville par rapport au paysage, le coucher du soleil contre la montagne en utilisant des indices globaux (par exemple, le spectre de puissance, informations sur l'histogramme des couleurs). L'étude de l'expérimentation de la classification des scènes naturelles passe par un ensemble de tâches multiples (détection, reconnaissance ou identification) et la stimulation (Réseaux, camouflage, images ambiguës, formes géométriques, etc.). Ces taches qui peuvent être utilisées dans les expériences sur la classification des scènes font appel à des niveaux de traitement très variés. Alors, dans cette étude, notre travail sera d'expérimenter une méthode de classification des scènes naturelles.

IV. Définition

Le terme de scènes naturelles se réfère à l'ensemble des images représentant le monde réel dans lequel on évolue et qui peuvent subir un changement d'état sous l'effet des actions des êtres vivants. Ces scènes qui peuvent être Intérieures ou Extérieures renferment des catégories ou classe d'image telles que : Scène Intérieure : Cuisine, Lit de maison, Salle à manger, Bureau, etc... Scène Extérieure : Jardin, Rue, Devanture de Maison, Forêt, Rivière, Mer, Paysage, etc...

V. Problèmes liés à ce domaine

La classification des scènes naturelles comme bon nombre d'autres du domaine de la reconnaissance de forme et de la vision par ordinateur fait face à des variables qui nuisent à ses performances. Ainsi, parmi les variables qui peuvent nuire à la perfection des systèmes de classification des scènes naturelles on peut citer :

• Variations de la luminosité dans les scènes
• Présence des images bruitées dans les scènes
• Présences des images floues dans les scènes
• Fond non uniforme des différents objets de la scène
• Forme géométrique ou non uniforme des objets de la scène
• Situation et relation avec l'espace

VI. Applications

Il trouve son domaine d'application dans divers domaines tels que la surveillance automatisée, la robotique, interaction homme machine, indication par vidéo et la navigation automobile. Annotation automatique de grandes bases de données d'images, vidéo, multimédia.

VII. Travaux Existants - Approches et Méthodes

L'état de l'art de la classification des scènes naturelles est marqué par différentes approches et méthodes de différents chercheurs qui ont proposé des techniques utilisées dans leurs travaux. Dans les lignes suivantes, nous faisons un tour d'horizon sur quelques différents travaux et articles afin de prendre connaissance de ces approches.

• Approches (Descripteurs et Classifieurs)

Différentes approches sont utilisées de nos jours pour expérimenter la classification des scènes naturelles. Nous pouvons par exemple citer, les approches basées sur les descripteurs comme :ACP, SIFT et Bag of Words.
Les approches basées sur les descripteurs locaux ET globaux, les histogrammes de gradients, les Filtres de Garbor sont entre autres les descripteurs les plus utilisés dans les approches de la classification des scènes naturelles y compris les classifieurs commme KNN, SVM et réseaux bayésiens.

• Dans un article intitulé, « Scene Recognition Based on Feature Learning from Multi-Scale Salient Regions» [4], les auteurs présentent une méthode efficace pour la reconnais-

sance de scène basée sur des fonctionnalités apprises à partir de régions saillantes à plusieurs échelles. La méthode trouve d'abord des régions saillantes multi-échelles dans une scène, puis extrait les fonctionnalités des régions via l'apprentissage par transfert en utilisant des réseaux de neurones convolutionnels (ConvNets). Les expériences sur deux ensembles de données de reconnaissance de scène populaires montrent que leur méthode proposée est efficace et a une bonne capacité de généralisation pour la reconnaissance de scène, par rapport aux benchmarks sur les deux ensembles de données. Cette méthode a affiché un taux de précision de 65.6% sur la base MIT-67.

• Dans un article publié récemment en Mars 2017 intitulé Scene classification of remote sensing images by optimizing visual vocabulary concerning scene label information, [1] L. Yan, Ruixi Zhu, Y. Liu, N. Mo ont proposé un algorithme de classification de scène d'image basé sur l'optimisation de mots visuels par rapport à l'information d'étiquette de scène pour traiter le problème du modèle traditionnel Bag Of Words (BOW) qui ne tient pas compte de l'information sur les étiquettes de scène des images de télédétection et de l'ambiguïté ou la redondance des vocabulaires visuels et qui n'est pas approprié aussi pour classer des antécédents similaires. La procédure d'algorithme est la suivante : Premièrement, les images sont divisées en patchs en utilisant la répartition spatiale des pyramides, puis les descripteurs (SIFT) sont extraits pour chaque image locale. Ces fonctionnalités sont ensuite regroupées avec K-means pour former un histogramme de chaque patch à différents niveaux en utilisant la stratégie de Boiman. Ils ont adopté « Image Frequency » comme méthode de sélection de descripteur des mots visuels dans chaque catégorie pour éliminer le vocabulaire visuel non pertinent pour une catégorie spécifique et obtenir un livre de codes spécifique à la classe. L'analyse en composante principale (ACP) est ensuite utilisée pour éliminer le vocabulaire visuel redondant. Cinq expériences ont été menées pour démontrer la performance de l'algorithme proposé se comporte mieux que d'autres méthodes représentatives dans les mêmes conditions. Ils ont totalisé une précision de 67% sur la base RSC11.

• Dans l'article intitulé, A Bayesian Hierarchical Model for Learning Natural Scene Categories [3], les auteurs ont proposé une approche pour apprendre et reconnaître les catégories de scènes naturelles. Ils ont représenté l'image d'une scène par une collection de régions locales, désigné comme code de mot obtenu par un apprentissage non supervisé. Dans cette approche, chaque région est représentée comme faisant partie d'un «thème». Leur algorithme fournit une approche fondée sur des principes pour l'apprentissage des représentations intermédiaires pertinentes des scènes automatiquement et sans supervision. L'approche présentée est un algorithme, un cadre probabiliste de principe pour apprendre des modèles de textures via des codes de mots (ou textons). Ces approches, qui utilisent des modèles d'histogramme de textons, sont un cas particulier de notre algorithme. Compte tenu de la flexibilité et la hiérarchie de notre modèle, de telles approches peuvent être facilement généralisé et étendu à l'aide de notre cadre. Leur modèle est capable de regrouper des catégories d'images en un sens hiérarchique, semblable à ce que les humains font. Ils ont totalisé une performance de 76% sur la base 13 scenes natural.

• MIT Scene Recognition
Une plateforme réalisée par les chercheurs du MIT dans lequel ils présentent une nouvelle base de données centrée sur la scène intitulée Places, avec 205 catégories de scène et 2,5 millions d'images avec une étiquette de catégorie. En utilisant le réseau neuronal convolutionnel (CNN), nous apprenons des fonctionnalités de scène profondes pour les tâches de reconnaissance de scène et établissons de nouvelles performances de pointe sur des repères centrés sur la scène. Nous fournissons ici la Base de données des lieux et les CNN formés à des fins

de recherche et d'éducation universitaire.[5]

VIII. Solution Proposée

• Approche : Descripteurs globaux et locaux

L'approche de notre solution proposée se repose sur les descripteurs globaux et locaux. Nous faisons une combinaison des descripteurs de fonctionnalités globales ainsi que des descripteurs de caractéristiques locales pour représenter chaque image, afin d'améliorer la précision de la reconnaissance. Nos étapes de base sont les suivantes.

• Descripteurs :
- Descripteur Global : Histogrammes des Gradients(HOG).
- Descripteur Local : Scale-Invariant Feature Transform (SIFT).

• Étape de Solution

Notre solution proposée est une combinaison des descripteurs locaux et globaux. Notre implémentation se fait suivant 4 étapes qui sont décrites ci-dessous.

1- Extraction des caractéristiques
Utilisation des descripteurs HOG et SIFT pour extraire les caractéristiques de chacune des images de la scène.

2- Encodage
Utilisation des fonctionnalités locales (SIFT) correspondant à chaque point clé, pour effectuer l'encodage. Nous utilisons ici le concept standard de «bag-of-visual-words». Nous qunatifions par la suite les caractéristiques dans les clusters et utilisons l'algorithme de K-means pour attribuer chaque mot visuel dans leur cluster.

3- Mise en commun
À cette étape, nous faisons une mise en commun des deux premières étapes. Ainsi, on fait une normalisation suivie d'une concaténation du descripteur global HOG correspondant à chaque image.

4- Clustering
La dernière étape, la plus importante de notre approche est la classification. Pour faire se faire, nous avons utilisé SVM comme classificateur. Pour implémenter SVM, nous avons utilisé la bibliothèque sckit-learn (sklearn).

IX. Implémentation

• Base d'image utilisée
Pour faire l'expérimentation, nous avons utilisé sur les données réelles de la base 13 Natural Scene Categories [6]

• Outils Utilisés
Nous avons programmé en Python, et utilisons les librairies comme OpenCV, Scipy, Skilearn, Numphy.

• Apprentissage
Pour faire l'apprentissage, nous avons utilisé les 100 premières images de chaque catégorie de la base.
• Test
Le test est réalisé sur toutes les autres images restantes dans les 13 catégories.
• Évaluation
Pour faire l'évaluation de notre système, nous avons utilisé plusieurs méthodes comme le Hold-Out et Matrice de Confusion. Ce qui donne en sortie du programme, la précision globale ainsi que la statistique standard de récupération d'informations telles que la précision, le rappel.

- Matrice de confusion
Représentation tabulaire spécifique qui contient les informations réelles et prédites effectuées par un système de classification. En colonne, on a les instances des classes prédites et en ligne les instances des classes réelles.

- Hold-Out
Cette méthode utilise une fonction d'optimisation maximisée qui permet de calculer le taux d'erreur, TE, taux de bon classement, TC avec TE = 100 – TC. D'où la formule de la précision pour le taux de bon classement TC se traduit par le nombre d'éléments bien classés sur le nombre total d'éléments. Voir les formules ci-dessous mentionnées :

$$\text{Précision} = \frac{\text{Nb éléments bien classés}}{\text{Nb éléments total}}$$

$$\text{Rappel}_i = \frac{\text{Nb éléments bien classés d'une classe } i}{\text{Nb éléments total dans la classe } i}$$

• Détails de l'Implémentation
Au lancement du programme, voici une liste d'action qu'exécute notre approche :
1- Lecture des repertoires, recupérations des différentes classes possibles et l'affection d'une classe à chaque individu d'un dossier(la classe étant simplement le nom du dossier).
2- Attribution des 100 premiers éléments de chaque dossier à l'ensemble d'entrainement le reste étant attribué au test.
3- Prétraitements sur les images(scale, changement d'échelle à une taille fixe et ouverture en niveau de gris).
3- Extraction des caractéristiques locales de chaque image(SIFT) et stockage dans une liste.
4- Application du K-means avec k prenant plusiseurs valeurs.
5- Ensuite formation des histogrammes en incrémentant de 1 selon l'appartenance de chaque point caractéristique aux différents clusters(on obtient un histogramme a k dimension, retenue pour décrire une image) et en même temps on calcule le descripteur global HOG correspondant à chaque image et on les garde.
6- Ici 03 choix s'offrent considerer seulement :
- Utilisser HOG comme un histogramme de k dimension pour décire les données test et d'entrainement.
- Utilsé SIFT comme descripteur pour décrire les données test et entraînement.
- Considérer les deux et les combiner.

Ainsi, nous avons combiné les deux et après entrainement du modèle SVM, on l'évalue et on trouve les résultats mentionnés ci-dessous.

X. Résultats et Analyse

Avec notre approche de combinaison des descripteurs globaux et locaux, nous avons un taux de précision égal à 81%. Le temps de calcul est de 25 minutes sur un core i3 intel 1.2 ghz.

Voir les diagrammes et captures ci-dessous : Dans le tableau ci-dessous, nous constatons la présence de nos 13 catégories de notre base. Au bas du tableau nous pouvons voir le taux de prédiction de 80.57%.

Fig.1 : Visualisation des différentes classes

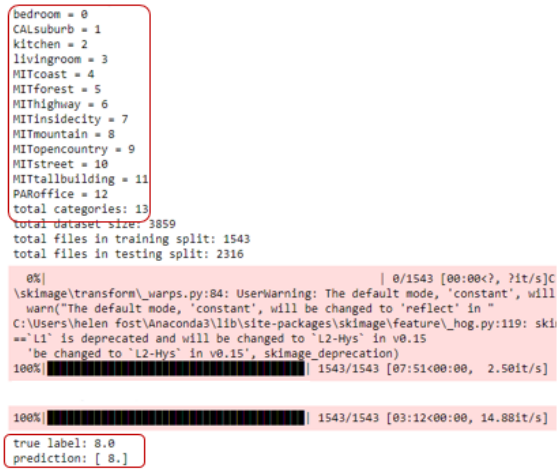
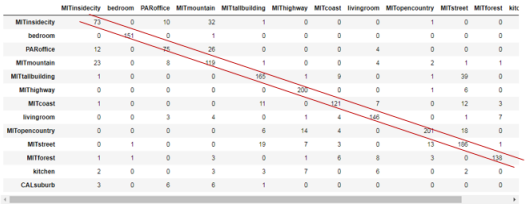


Fig2 : Matrice de confusion



L'encadré en rouge représente le nombre d'élément bien classés pour chaque catégorie ou classe.

Fig3 : Précision et rappel du système

```
No. of test instances: 2316 2316
Overall accuracy: 0.805267702936
                precision    recall

       bedroom       0.63      0.58
     CALsuburb       0.99      0.99
       kitchen       0.74      0.58
    livingroom       0.61      0.70
      MITcoast       0.80      0.76
     MITforest       0.87      0.96
    MIThighway       0.82      0.78
  MITinsidecity       0.83      0.80
   MITmountain       0.91      0.82
 MITopencountry       0.70      0.81
    MITstreet       0.92      0.80
MITtallbuilding       0.83      0.88
     PARoffice       0.80      0.86

   avg / total       0.81      0.81
```

Dans le tableau ci-dessus, on peut constater le taux de précision global du système, 81Nous pouvons aussi visualiser le taux de bon classement et le rappel pour chaque catégorie. La classe living room avec le taux de classement le plus faible.
la classe Calsuburb obtient le meilleur taux de détection : 99 %, ici le descripteur global utilisé nous permet de mieux détecter les maisons. La classe LivingRoom obtient le pire taux de détection : 61 % malgré l'utilisation des deux descripteurs, il reste toujours plusieurs confusions entre livingroom et bedroom

XI. Comparaison avec la littérature

Notre approche consiste en une combinaison des descripteurs globaux et locaux. Alors, pour comparer notre méthode à la littérature, nous avions implémenté de façon détaillée une méthode pour chacun des descripteurs que nous avions combinés, HOG et SIFT.

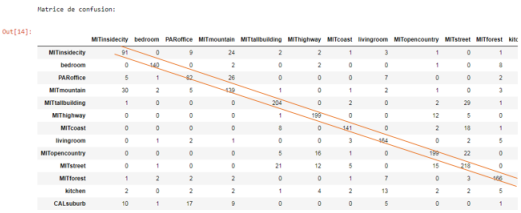
• Approche Simple - Histogramme des gradients - HOG

Avec l'approche de l'Histogramme des gradients seulement, (HOG), nous obtenons un taux de précision de 79%. Le temps de calcul est d'environ de 13 minutes sur un core i3 intel 1.2 ghz.

Fig. 4 : Visualaisation des différentes catégories

```
bedroom = 0
CALsuburb = 1
kitchen = 2
livingroom = 3
MITcoast = 4
MITforest = 5
MIThighway = 6
MITinsidecity = 7
MITmountain = 8
MITopencountry = 9
MITstreet = 10
MITtallbuilding = 11
PARoffice = 12
total categories: 13
Taille totale du jeu de données: 3859
Total des fichiers du jeu de données : 3859
Total des labels dans le jeu de données: 3859
Nombre total des fichiers du split entrainement: 1300
Nombre total des fichiers du split de test: 2559
...Chargement du fichier de kmean...
...Chargement du fichier apprentissage...
true label: 7.0
prediction: [ 11.]
```

Fig. 5 : Matrice de confusion



L'encadré représente le nombre d'élément bien classés pour chaque catégorie ou classe.

Fig. 6 : Précision et rappel du système

```
Nombre instances test: 2559 2559
Précision globale:
0.794060179758
                precision    recall

       bedroom       0.65      0.64
     CALsuburb       0.95      0.88
       kitchen       0.69      0.59
    livingroom       0.68      0.72
      MITcoast       0.84      0.85
     MITforest       0.85      0.92
    MIThighway       0.90      0.82
  MITinsidecity       0.82      0.80
   MITmountain       0.85      0.80
 MITopencountry       0.73      0.80
    MITstreet       0.86      0.86
MITtallbuilding       0.74      0.83
     PARoffice       0.79      0.65

   avg / total       0.80      0.79
```

- Analyse

La classe Calsuburb obtient le meilleur taux de détection : 95%. Ici l'utilisation de HOG comme descripteur permet de mieux détecter les maisons car globalement elles ont la même forme, avec leur base à peinture claire et leur toiture toujours foncée facile à détecter avec HOG.

La classe Bedroom obtient le pire taux de détection : 65%. La raison c'est que globalement en utilisant HOG, les objetcs de la classe Bedroom ressemblent beaucoup à ceux de la

classe livingRoom. Ainsi le modèle confond entre ces deux classes. Il fallait ajouter un descripteur local pour améliorer cette détection afin de mieux distinguer entre ces deux classes.

• Approche Simple - SIFT + BOW

Avec l'implémentation du Bags of Words (BOW) avec les points caractéristiques locaux (SIFT) comme descripteurs pour entrainer le modèle SVM, nous obtenons un taux de précision de 75%. Le temps de calcul est alors d'environ 15 minutes sur un core i3 intel 1.2 ghz.

Fig. 7 : Visualaisation des différentes catégories

```
Le jeu de données est déjà téléchargé et extrait!
bedroom = 0
CALsuburb = 1
kitchen = 2
livingroom = 3
MITcoast = 4
MITforest = 5
MIThighway = 6
MITinsidecity = 7
MITmountain = 8
MITopencountry = 9
MITstreet = 10
MITtallbuilding = 11
PARoffice = 12
total categories: 13
Taille totale du jeu de données: 3859
Total des fichiers du jeu de données : 3859
Total des labels dans le jeu de données: 3859
Nombre total des fichiers du split entrainement: 1300
Nombre total des fichiers du split de test: 2559
```

Fig. 8 : Matrice de confusion

Out[12]:

	MITinsidecity	bedroom	PARoffice	MITmountain	MITtallbuilding	MIThighway	MITcoast	livingroom	MITopencountry	MITstreet	MITforest	kitc
MITinsidecity	62	1	16	38	1	0	0	3	2	1	2	
bedroom	1	150	0	4	0	0	0	1	0	1	2	
PARoffice	21	0	72	30	0	0	0	7	0	0	1	
MITmountain	31	1	29	104	1	0	0	6	1	1	0	
MITtallbuilding	0	0	0	0	206	0	8	0	1	25	0	
MIThighway	0	1	0	0	1	196	9	0	8	7	2	
MITcoast	0	1	0	1	10	2	139	1	2	9	3	
livingroom	0	0	0	1	4	3	5	148	0	5	16	
MITopencountry	1	0	0	1	9	10	3	0	202	19	3	
MITstreet	0	1	0	0	33	16	6	1	16	197	1	
MITforest	0	5	1	2	1	1	6	12	5	2	153	
kitchen	6	0	0	10	3	6	0	10	4	2	1	
CALsuburb	5	0	12	23	0	0	0	0	0	0	0	

L'encadré représente le nombre d'élément bien classés pour chaque catégorie ou classe.

Fig. 9 : Précision et rappel du système

```
Nombre instances test: 2559 2559
Précision globale:
 0.752637749121
                precision    recall

       bedroom       0.49      0.43
     CALsuburb       0.94      0.94
       kitchen       0.52      0.52
    livingroom       0.49      0.54
      MITcoast       0.76      0.86
     MITforest       0.84      0.91
   MIThighway       0.83      0.81
MITinsidecity       0.77      0.73
   MITmountain       0.84      0.81
MITopencountry       0.73      0.72
     MITstreet       0.83      0.79
MITtallbuilding       0.87      0.82
     PARoffice       0.74      0.71

avg / total       0.75      0.75
```

- Analyse

La classe Calsuburb obtient le meilleur taux de détection : 94%. La raison ici est que le descripteur local utilise les caractéristiques des immeubles, or les immeubles ne contiennent pas d'autres types d'objets c'est pour cette raison que ce type d'objet est mieux détecté avec le descripteur local.

La classe Bedroom et livinroom obtiennent le pire taux de détection : 49% . En effet, dans l'ensemble d'entraînement on observe une grande diversité d'objets (formes de lits, de meubles très variés) présents dans la chambre et le salon à des différentes positions, c'est pour cela que le descripteur local ne s'ensort pas bien à ce niveau, la position des objets (chaise, table, lit mis à des positions très différentes d'une image à l'autre)

XII. Discussion

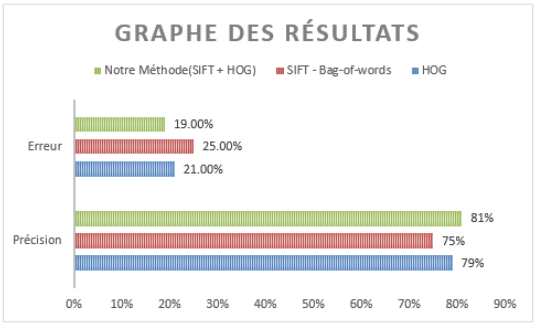
D'une manière générale, pour chacune des solutions proposées, l'approche de notre solution proposée (combinaison de HOG et SIFT) et l'implémentation séparée des HOG et SIFT, la seule couche qui change est celle du descripteur pour entrainer le classifieur SVM.

Par rapport aux résultats obtenus, nous avons constaté que notre solution, qui est une combinaision des descripteurs globaux et locaux affiche un meilleur taux de précision par rapport à l'implémentation détaillée de SIFT et HOH.
Une raison à cette forte perfomance de la combinaison des deux c'est que les descripteurs locaux détectent parfois certains traits ou aspects là où les descripteurs globaux n'arrivent pas.
D'où, en terme de précision, la combinaison des deux descripteurs vaut mieux. Le tableau et diagramme ci-dessous permettent d'observer ces valeurs de précisions.

Fig. 10 : Tableau de comparaion

Méthode	Précision	Erreur	Temps
HOG	79%	21.00%	13 min
SIFT - Bag-of-words	75%	25.00%	15 min
Notre Méthode(SIFT + HOG)	81%	19.00%	25 min

Fig. 11 : Diagramme de comparaion des taux



En terme de temps de calcul, l'implémentation détaillée des deux descripteurs valent mieux que notre méthode (Fig 10). La raison c'est que en combinant les deux méhodes, le temps d'apprentissage de notre modèle augmente normalement.

XIII. Conclusion

Le problème de classification des scènes naturelles requiert beaucoup d'attention dans la recherche. Plusieurs auteurs ont tenté de proposer de multiples approches pour des applications pratiques.

Les méthodes utilisées pour la mise en oeuvre de la classification des scènes naturelles sont diverses et se basent pour certaines directement sur les caractéristiques des images dans la scène et pour d'autre sur les algorithmes d'apprentissage ou encore des approches statistiques.

Dans ce rapport, nous avons présenté l'implémentation et les résultats de notre solution basée sur la combinaison des descripteurs globaux et locaux.
Nous avons fait une approce en quatre étapes pour notre solution. Comme descripteurs globaux nous allons utilisé les Histogrammes de Gradiant. Comme descripteurs locaux, nous avons utilisé SIFT. Comme classifieur, nous avons proposé d'utiliser SVM.
L'**avantage**de notre approche réside dans la robustesse de la combinaison des deux descripteurs qui nous donnent ce taux de 81%.
Mais, notre solution a l'**inconvénient** d'être coûteuse en terme de temps et de mémoire.

Références

[1] L. Yan, Ruixi Zhu, Y. Liu, N. Mo, Scene classification of remote sensing images by optimizing visual vocabulary concerning scene label information,Mars 2017.
[2] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. DeCAF : A deep convolutional activation feature for generic visual recognition. In International Conference on Machine Learning (ICML), 2014
[3] A Bayesian Hierarchical Model for. Learning Natural Scene. Categories. L. Fei-Fei and P. Perona. CVPR 2005. Presented By. N. Soumya, ME (SSA), 2017.
[4] Scene Recognition Based on Feature Learning from Multi-Scale Salient Regions Dianzi Keji Daxue Xuebao/Journal of the University of Electronic Science and Technology of China 46(3) :600-605 • March 2017
[5] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. "Learning Deep Features for Scene Recognition using Places Database." Advances in Neural Information Processing Systems 27 (NIPS), 2014
[6] 13 Natural Scene Categories) http ://vision.stanford.edu/resourceslinks.html