



LEMOS Lucas
Informatique et Electronique des Systèmes Embarqués
POLYTECH GRENOBLE
Rapport de stage IESE4

**Modèle bio-inspiré de perception visuelle pour la
catégorisation d'images de scènes naturelles**

Tome principal

2016/2017
Stage du 02 mai au 28 juillet 2017

Table des matières

1	Le laboratoire	2
1.1	Présentation générale	2
1.2	Ma position dans le laboratoire	3
2	Introduction	3
3	Gestion du projet	4
4	Modèle de perception visuelle bio-inspiré	5
4.1	Filtre Spatialement Variant	6
4.2	Rétine	7
4.3	Normalisation 1	8
4.4	Banc de filtres de Gabor	8
4.5	Normalisation Fréquence	11
4.6	Normalisation 2	12
5	La base d'images	12
6	Algorithmes de classification	14
7	Résultats	14
7.1	Analyse des paramètres du banc de filtres de Gabor	16
7.1.1	Largeur de bande radiale	17
7.1.2	Largeur de bande transversale	18
7.2	Approche “ <i>imagette</i> ”	19
7.3	Classification	20
7.3.1	Validation Croisée	20
7.3.2	Réseau Neurone	22
8	Conclusion	24
	Acronymes	25

1 Le laboratoire

1.1 Présentation générale

Le stage s'est déroulé dans un laboratoire de recherche qui s'appelle GIPSA-lab (Grenoble Images Paroles Signal Automatique). L'acronyme révèle les 3 départements qui composent le noyau des thématiques auxquelles le laboratoire s'intéresse. La liste ci-dessous détaille les objectifs scientifiques de chaque département.

1. *Automatique* : Développement des outils d'analyse, de modélisation, de diagnostic et de commande de systèmes dynamiques, embarqués, complexes et en réseau. Inscription des nouvelles méthodologies dans les enjeux applicatifs et leur validation par des implémentations expérimentales.
2. *Parole - Cognition* : Progression des connaissances sur la parole, le langage et la communication parlée. L'étude et modélisation des signaux de la parole, les systèmes physiologiques et cognitifs qui les produisent ainsi que les systèmes linguistiques qui les organisent. Création des agents conversationnels virtuels capables de communiquer avec des interlocuteurs humains.
3. *Images - Signal* : Développement des modèles avancés sur les images et les signaux en puisant les problématiques dans les systèmes physiques, biologiques ou artificiels. Conception des algorithmes de traitement performants. Validation des modèles dans des cadres applicatifs réels.

Chaque département est sous-divisé en équipes de recherche (voir organigramme de la figure 1).

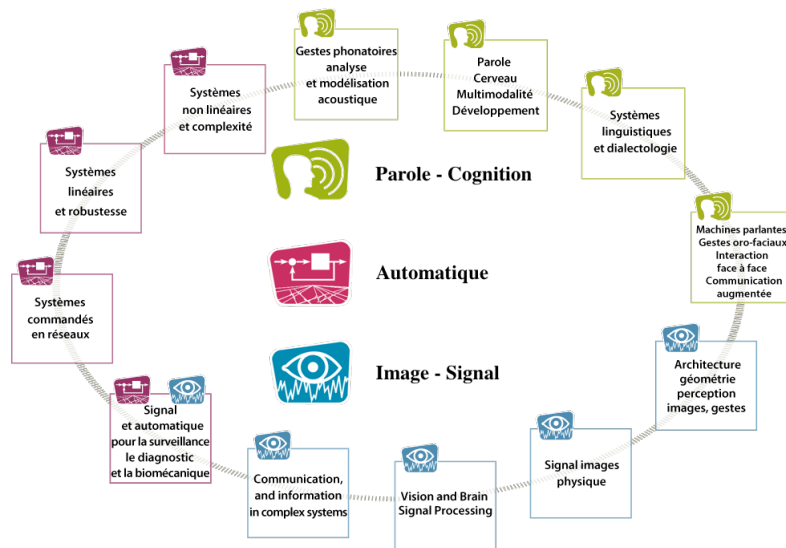


Figure 1: Les 12 équipes de recherches qui composent les 3 départements du GIPSA-lab.

Source: <http://www.gipsa-lab.fr/>

1.2 Ma position dans le laboratoire

J'étais encadré par madame Nathalie Guyader dans le département images et signal, dans l'équipe Vision and Brain Signal Processing (ViBS). Les thématiques de notre équipe sont les suivantes :

- Méthodologie d'analyse des signaux cérébraux
- Modèles fonctionnelles et cognitifs de perception visuelle
- Interface cerveau-machine et neurofeedback

Comme je l'expliquerai plus tard, mon stage a porté sur le développement d'un modèle computationnel de la perception visuelle de l'homme pour la catégorisation d'images de scènes naturelles (images de plages, de villes, de montagnes, de forêts, de scènes d'intérieurs etc.)

Le GIPSA-lab comprend 350 membres (chercheurs, doctorants, administratifs, post-doctorants et stagiaires). Le laboratoire a environ 500 publications par an et un budget de 4 million d'euros.

Les domaines d'applications sont très variés : la santé, l'environnement, l'énergie, la géophysique, les systèmes embarqués, la mécanique, entre autres. En prenant ça en compte et aussi le fait le laboratoire n'a pas un produit finale, dont la sûreté de fonctionnement forcément demanderait une norme ISO, GIPSA-lab développe des projets qui visent l'efficacité de consommation énergétique, bien comme des sources alternatifs ce qui fait qu'il ait une vision sur la thématique du développement durable.

2 Introduction

Les 5 sens (goût, odorat, vue, ouïe, toucher) sont nos principaux outils de communication verbale et non verbale. C'est grâce à eux que nous sommes en communication et en relation avec l'environnement et les autres. Dans le cadre de mon stage je me suis particulièrement intéressé à la vue. Le but de mon stage est de développer un modèle computationnel de perception visuelle qui simule les différentes étapes de traitement de l'information visuelle depuis la rétine jusqu'aux cellules corticales. Le but de mon travail était plus précisément de reprendre des programmes existants, de les uniformiser et surtout de les assembler en testant les différentes parties. Nous avons décidé en début de stage qu'il fallait pour effectuer ces tests se baser sur une application et donc des mesures de performances. Nous avons choisi de tester notre modèle sur une tâche de classification d'images. Ainsi notre modèle qui aboutit à une description d'image inspirée de la perception visuelle devrait permettre de catégoriser selon un algorithme de classification "basique" les grandes classes que nous avons dans nos images, à savoir des images de scènes naturelles (paysages) vs des images de scènes artificielles (contenant des objets maisons, constructions humaines etc).

Différentes études montrent qu'on est capable de classifier des images complexes juste en jetant un coup d'oeil grâce aux propriétés spatiales de la scène, ce qui permet la classification en catégories sémantiquement abstraites comme naturelle, artificielle, ouverte, fermé [1], [2], [3]. Notre modèle profite de cette conclusion pour extraire les caractéristiques spatiales d'une image et la classifie soit comme naturelle, soit comme artificielle.

Le projet, donc, se déroule autour de la thématique de la perception visuelle des êtres humains, où l'objectif principal est de quantifier l'impact des différents paramètres du modèle sur les pourcentages de bonne classification des images. On commence par la prise en main du modèle déjà existant dans l'équipe. Ce modèle est développé en Matlab. Il comprend différentes étapes (décrites à la section 4). Ainsi dans ce rapport je présente les objectifs et la gestion de projet que j'ai suivie durant mon stage, le modèle existant et les tests que j'ai réalisés pour la validation des différents paramètres ainsi qu'une discussion sur les résultats obtenus.

3 Gestion du projet

J'ai commencé le stage en 21/05 et je l'ai terminé en 28/07, ce qui donne exactement 12 semaines de travail. Pour accomplir les objectifs mentionnés précédemment, le travail a été divisé selon les informations de la table 1.

Table 1: Chronogramme du travail réalisé pendant les 12 semaines de stage.

Semaines	Activité
1	Étude bibliographique – thèse N. Guyader (PhD thesis, 2004) + les travaux (codes et notes) de Ralucca Vlad Post doctorante au GIPSA-lab
2	Travail pratique sur le filtre de Gabor
3	Mise en pratique des différents algorithmes de traitement de données multidimensionnelles : Analyse en Composantes Principales et Analyse Discriminante Linéaire.
4 - 5	Choix du classifieur linéaire et tests du modèle déjà existant au laboratoire sur différentes bases d'images de scènes naturelles
6 - 8	Tests de l'influence de différents paramètres du modèle sur le taux de bonne classification
9 - 12	Test préliminaires de l'approche imagerie + validation des résultats avec une validation croisée + test du modèle en utilisant un réseau neurone comme classifieur

4 Modèle de perception visuelle bio-inspiré

Le modèle de description d'image utilisé ici repart d'un modèle initialement proposé par N. Guyader (PhD thesis, 2004 [4]). Ce modèle s'inspire du fonctionnement du système visuel humain pour proposer une chaîne de traitement d'images. Une image sera à l'issue de ce modèle décrite (ou représentée) par quelques valeurs qui correspondent à l'énergie de l'image filtrée par un banc de filtres de Gabor. Dans sa thèse, N Guyader a modélisé les différentes étapes depuis la réception de l'information lumineuse par les photorécepteurs de la rétine jusqu'aux traitements des cellules corticales des aires visuelles primaires. Le modèle proposé comporte 6 étapes de traitement ; un schéma en est donné à la figure 2.

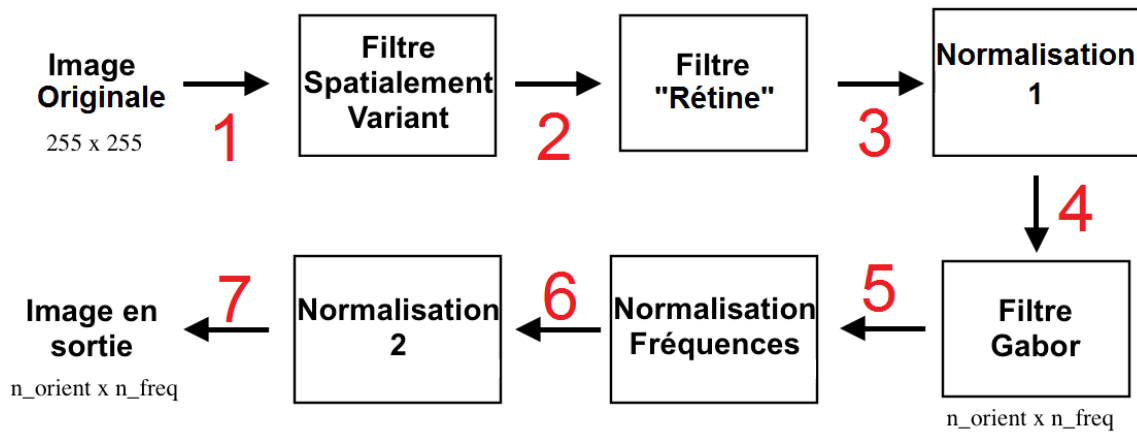


Figure 2: Simulation des étapes de traitement de l'image par le modèle de perception visuelle inspiré de la perception visuelle chez l'homme. Les numéros indiqués en rouge correspondent aux différentes étapes de traitement et sont par la suite repris et détaillés

La mise en oeuvre du modèle est déjà faite sur *Matlab*. On travaille avec des images fournies par SUN qui peuvent être accédées en cliquant [ici](#). Les images ont 255 x 255 pixels et sont en noir et blanc. L'idée d'utiliser l'image en niveaux de gris c'est qu'on gagne en rapidité et aussi que, selon un travail qui n'a pas encore été publié, la couleur n'influence pas lors de la catégorisation de scènes naturelles vs. artificielles.

A mon arrivé, le modèle était testé sur une application de catégorisation des scènes : artificielles vs. naturelles et l'algorithme utilisé pour la classification était un Support Vector Machine (SVM) [5]. Comme mon objectif c'était de valider l'impact des différentes étapes de traitement, on a décidé d'implémenter un classifieur plus simple afin de tester plus le modèle de description des images (et donc le modèle bio-inspiré) que le modèle de classification utilisé. Remarque : l'étape 4 qui consiste au filtrage de l'image par un banc de filtres de Gabor simulant les réponses des cellules du cortex visuel primaire est l'étape qui contient le plus grand nombre de

paramètres et celle sur laquelle j'ai beaucoup travaillé et effectué de nombreux tests. Dans cette étape, l'image originale (255 x 255 pixels) est décrite par une matrice de taille $n_orient \times n_freq$ éléments, ce qui est détaillé dans la section 4.4. Chaque élément correspond à l'énergie de l'image en sortie d'un filtre de Gabor accordé sur une certaine fréquence spatiale et sur une certaine orientation.

On illustre, maintenant, le fonctionnement de chaque étape sur une image donnée. La figure 3 montre une exemple d'une image artificielle et naturelle. Ces deux images sont très classiques et, donc, plus simple à classer. La simplicité de la tâche provient des orientations contenues dans les images qui sont bien différentes entre l'image artificielle (figure 3a) et l'image naturelle (figure 3b).



(a) Image artificielle



(b) Image naturelle

Figure 3: Exemple des deux types d'images à catégoriser

Les figures 4 et 7 montrent les transformations de l'image 3a dans chaque étape du modèle. Le nombre en dessous de chaque image représente les étapes du traitement, en utilisant la numérotation des étapes du modèle donnée au schéma de la figure 2.

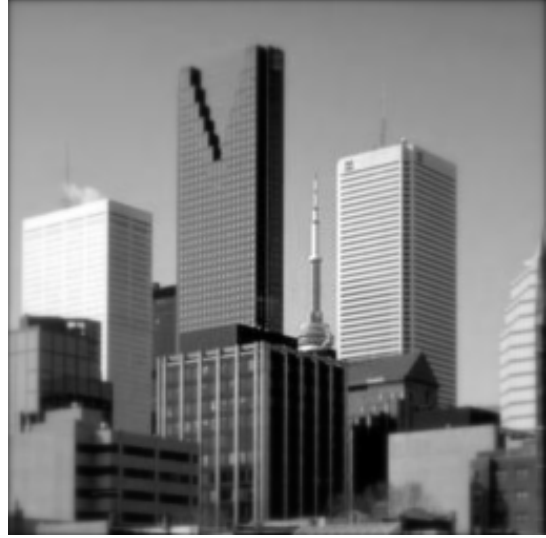
4.1 Filtre Spatialement Variant

Tout d'abord, l'image d'entrée passe par le filtre spatialement variant (Étape 2). Il joue le rôle de l'acuité visuelle qui, chez nous, les êtres humains, décroît avec l'excentricité. Ainsi la région que nous regardons est nette, mais les régions qui se situent autour de la région regardée ont une acuité visuelle moindre, c'est-à-dire qu'il est plus difficile d'en percevoir les détails. Nous pouvons ainsi facilement lire des lettres en vision centrale mais cette lecture est bien plus difficile voir impossible en vision périphérique. Le filtre spatialement variant va créer, à partir du centre de l'image (on suppose que les images sont perçues en vision centrale : le regard serait

alors au centre des images), des cercles concentriques et, au fur et à mesure que les rayons augmentent, l'image devient moins nette (voir image de sortie de l'étape 2 - figure 4b).



(a) Étape 1



(b) Étape 2



(c) Étape 3



(d) Étape 4

Figure 4: Quatre premières étapes du modèle de perception visuelle. Étape 1 : Image originale. Étape 2 : Image après le filtre spatialement variant. Étape 3 : Image après le filtre "rétine". Étape 4 : Image après la normalisation 1.

4.2 Rétine

En sortie du filtre spatialement variant, l'image est pré-traitée par l'étape de "Rétine". Ainsi comme indiqué par N.Guyader [4] :

L'un des rôles de la rétine est de permettre le codage des images, c'est-à-dire, la captation de l'intensité lumineuse, malgré des niveaux d'éclairement très variés... Grâce à la présence de deux types de photorécepteurs et au phénomène d'adaptation que le système visuel fonctionne de manière optimale.

L'idée, en utilisant les cellules parvocellulaires comme sortie, c'est de ressortir les variations locales de contraste et les détails.

4.3 Normalisation 1

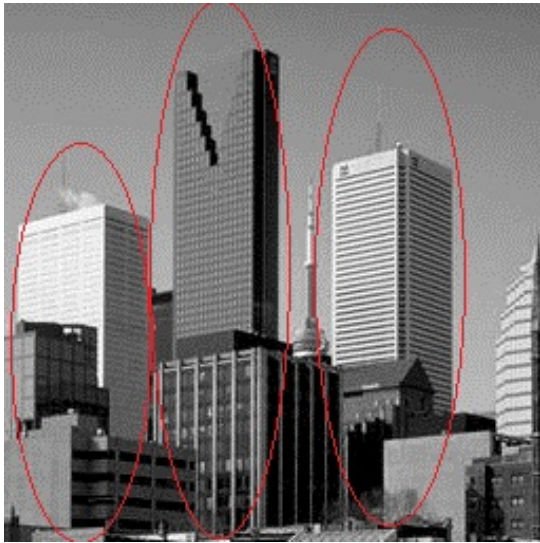
Ensuite, une normalisation de la luminance des pixels de l'image est réalisée (Étape 4). Pour chaque image, la luminance de chaque pixel est centrée sur la luminance moyenne de l'image et divisée par l'écart type de la luminance sur l'image (on centre et on réduit la luminance des pixels d'une image à une moyenne de 0 et un écart type de 1).

4.4 Banc de filtres de Gabor

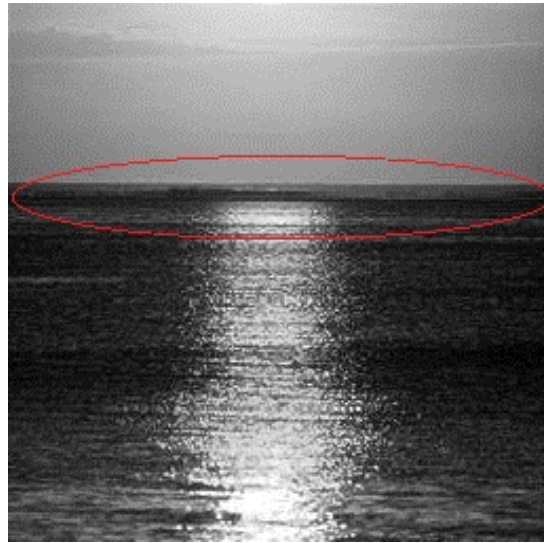
Après les étapes de traitement qui simulent le fonctionnement de la rétine, l'image est décomposée par un banc de filtres de Gabor qui simule le fonctionnement de certaines cellules du cortex visuel primaire. Le bloc qui contient le plus de paramètres est celui du banc de filtres de Gabor c'est pour cela que ce bloc est décrit en détails. On commence par la transformée de Fourier (TF) d'une image. Cette opération nous permet de récupérer les fréquences spatiales et les orientations présentes dans une image. On représente dans la figure 5, la TF des deux images utilisées comme exemples - figure 3 - et aussi les images originales pour faciliter la visualisation.

On observe que dans le cas de l'image artificielle, figure 5a, le module, avec un affichage logarithmique, de sa TF, figure 5c, montre, en gros, trois lignes horizontales en blanc. Le plus claire, le plus fortes sont les composantes. On regarde que ces trois lignes représentent les trois orientations verticales prédominantes, remarqués par les traces en rouges. Le même raisonnement peut être fait pour les figures 5b et 5d. L'orientation horizontale dans la mer apparaît comme une ligne vertical dans sa TF.

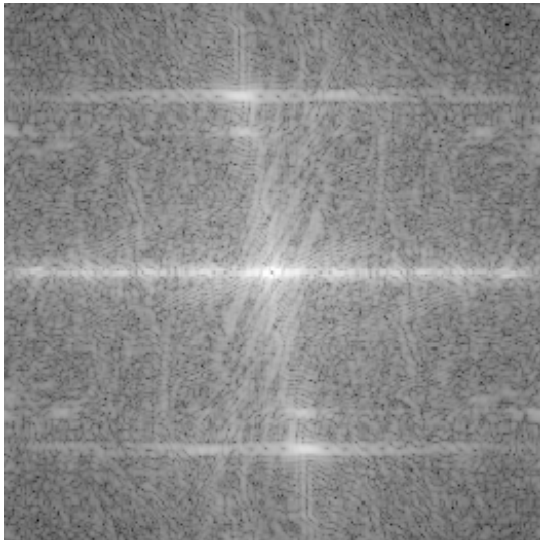
Un filtre de Gabor est un filtre passe-bande orienté qui permet de récupérer dans une image son énergie autour d'une certaine fréquence spatiale et selon une orientation. Ainsi, si on a m fréquences et n orientations, on peut représenter une image par $m \times n$ valeurs correspondantes aux énergies de l'image à ces fréquences et orientations. Si on arrive, donc, à bien choisir les valeurs de m et n , on est capable de, à partir du filtre de Gabor, prendre une décision sur quel group une tel image appartient. La figure 6 montre deux exemples d'un banc de filtres de Gabor, le premier avec 4 orientations et 3 fréquences, et le dernier avec 6 orientations et 5 fréquences.



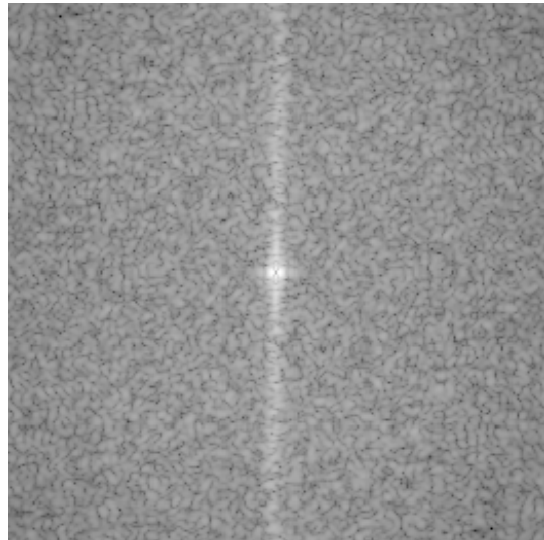
(a) Image artificielle



(b) Image naturelle



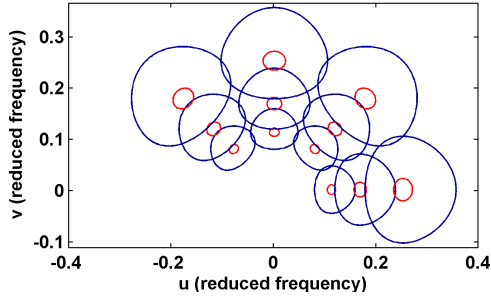
(c) Logarithme du module de la TF de l'image artificielle



(d) Logarithme du module de la TF de l'image naturelle

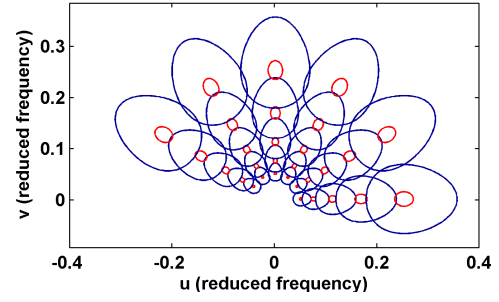
Figure 5: Les deux images exemples et le logarithme des modules des TF. Le centre de l'image correspond à la fréquence nulle et les hautes fréquences spatiales sont en périphérie. Le blanc correspond à une forte concentration d'énergie

Half-height contour of the 4x3 Gabor filter bank



(a) Banc de filtre de Gabor 4x3

Half-height contour of the 6x5 Gabor filter bank



(b) Banc de filtre de Gabor 6x5

Figure 6: Exemple de banc de filtres de Gabor. Visualisation des filtres à mi-hauteur dans le plan des fréquences spatiales réduites. La fréquence nulle étant au centre de la représentation

La figure 7 montre le résultat du filtre de Gabor - étape 5 - de l'image artificielle - figure 3a. On a choisi 6 orientations de 0° à 150° , selon l'équation 4.1, et 5 fréquences différentes. La première orientation est toujours 0° , $\theta_0 = 0^\circ$. Une fréquence spatiale de 0,5 signifie qu'un pixel sur 2 change de luminance soit ici pour une image de 255 pixels. Dans notre cas, on définit une fréquence maximale de 0,25 et les autres quatre fréquences sont calculées selon l'équation 4.2.

$$\theta_{i+1} = \theta_i + \frac{180^\circ}{\text{no_orientations}} \quad (4.1)$$

$$f_k = \frac{f_{max}}{1,5^k} \quad (4.2)$$

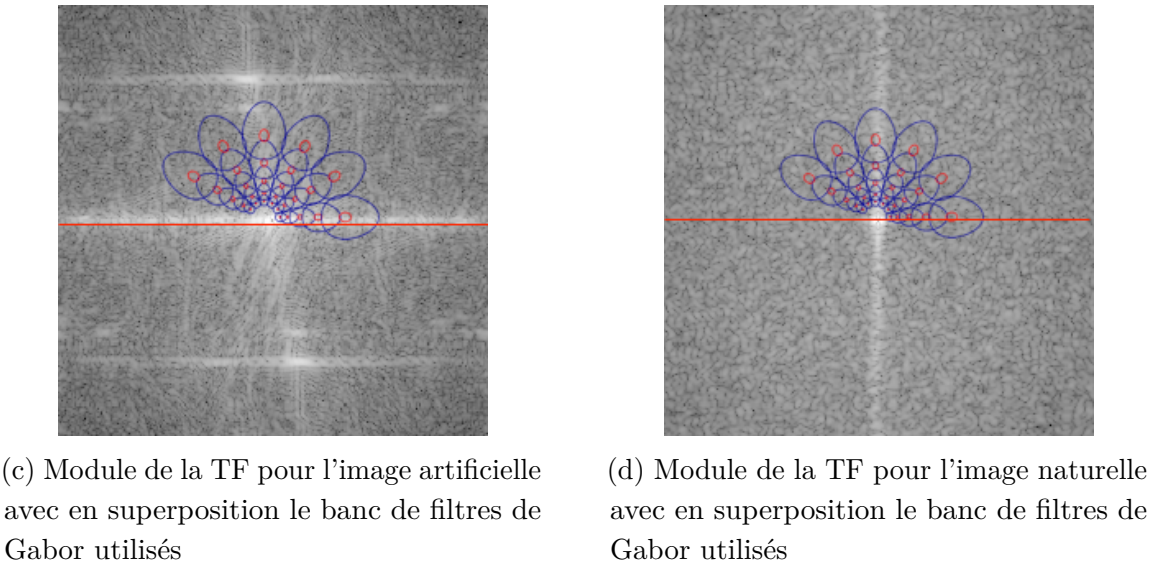
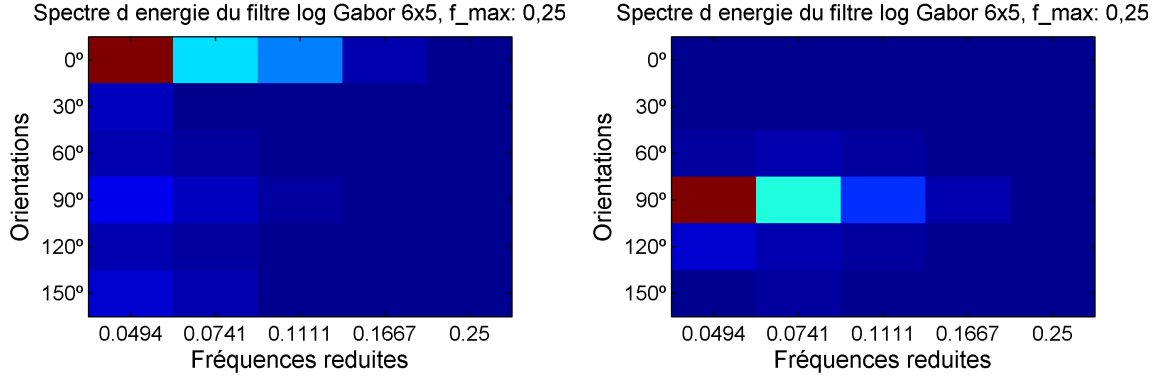


Figure 7: L'étape 5 du modèle de perception visuelle pour les deux images exemples.

On voit bien que l'énergie, pour l'image naturelle - figure 7b, est bien autour de 90° . Remarque : l'orientation de la TF est orthogonale à celle du signal spatial. Pour le cas de l'image artificielle - figure 7a, on regarde que l'angle prédominant c'est le 0° . Les figures 7c et 7d illustrent le module de la TF pour les deux images avec des axes référentiels, en rouge, pour l'orientation. Donc, les énergies dans les angles choisis sont mesurés entre les lignes blancs et la rouge. On voit aussi le banc de filtre de Gabor sur les modules des TF, il nous donne une idée de quels sont les tranches qui sont prise en compte pour constituer les matrices dans les figures 7a et 7b.

4.5 Normalisation Fréquence

La normalisation fréquence - étape 6 - est faite pour que les énergies, dans chaque fréquence, se retrouvent entre 0 et 1. La figure 8 montre cet effet dans la matrice en sortie du banc de filtres de Gabor.

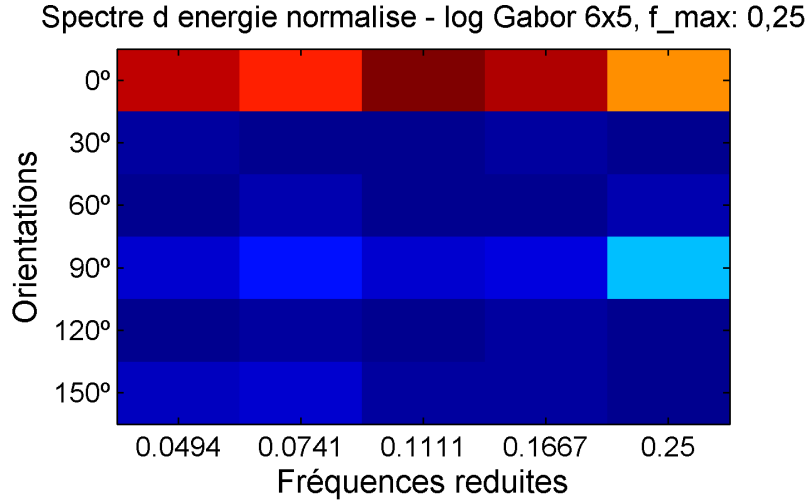


Figure 8: Étape 6 du modèle de perception visuelle. Chaque élément de la matrice correspond à l'énergie de l'image d'entrée filtrée par le filtre de Gabor à une certaine fréquence et une certaine orientation.

4.6 Normalisation 2

La normalisation 2 - étape 7 - était mise en oeuvre pour faciliter la classification avec SVM. Elle range les valeurs aussi entre 0 et 1. Les valeurs de la matrice de Gabor, néanmoins, restent distribuées d'une façon similaire et, les valeurs étant simplement normalisées leur visualisation par une image ne diffère pas de l'image de sortie de l'étape 6.

5 La base d'images

On utilise une base de données principale, qu'on appelle *training*. Elle possède 1600 images, dont 800 sont naturelles et 800 artificielles. On a aussi la base de *Grenoble*, qui sont 200 images prélevé du *training*. Ces images sont les plus prototypiques des images naturelles et des images artificielles.. La dernière base de donnée est la base de *Lille*, où les images sont choisit de façon à celle de *Grenoble*, ce sont les images les plus durs à classer. Elle a aussi 200 images, dont 100 sont naturelles et 100 artificielles. La figure 9 montre des exemples des images naturelles et artificielles de chaque base de données.



(a) Image naturelle - *dataset* Grenoble



(b) Image artificielle - *dataset* Grenoble



(c) Image naturelle - *dataset* Lille



(d) Image artificielle - *dataset* Lille



(e) Image naturelle - *dataset training*



(f) Image artificielle - *dataset training*

Figure 9: Exemple des images des trois *datasets*. Le *dataset* principal : *dataset training* contenant 1600 images, le *dataset* Grenoble contenant les 200 images les plus prototypiques et le *dataset* Lille contenant des images tirées aléatoirement du *dataset training*

6 Algorithmes de classification

Le but est ici de chercher si des classifieurs simples permettent d'obtenir des performances similaires à celles obtenues par le SVM. Il sera ensuite plus aisé d'effectuer l'ensemble de nos tests avec ce classifieur simple. La première démarche c'était d'utiliser la méthode Linear Discriminant Analysis (LDA) [6]. Cette méthode a été vu dans le cours de Traitement des Données et il essaye de trouver l'axe qui distingue le plus les données mis à son entrée. Pour le faire, il faut lui dire à quel groupe chaque donnée appartient. La figure 10 montre une exemple. On lui donne 100 images de scènes naturelles et 100 images de scènes artificielles. Dans ce cas de figure, on voit que l'axe discriminant 1 sépare bien les deux types d'images.

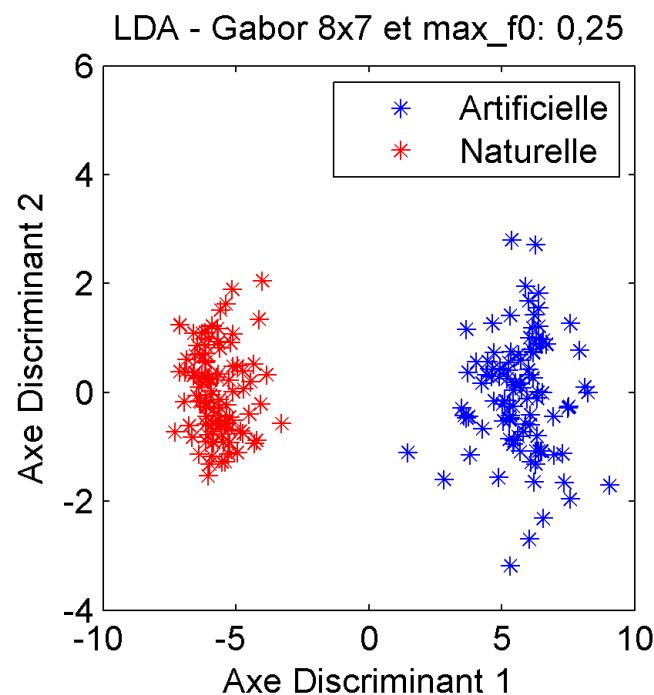


Figure 10: Méthode LDA appliqué après une filtrage de Gabor avec 8 orientations, 7 fréquences, dont la maximale est 0,25

En ayant ce résultat en tête, on a décidé de garder cet classifieur comme le défaut pour qu'on puisse jouer sur les paramètres d'autres blocs du modèle

7 Résultats

Ayant fixé un classifieur simple nous l'avons utilisé pour tester les différents paramètres de notre banc de filtres de Gabor. Ainsi nous utiliserons par la suite les paramètres du banc de filtres qui permettent d'obtenir des résultats de classification optimaux sur les différentes *datasets*. En regardant la TF des images, on a réparé que

l'énergie des images allaient jusqu'à 0,3 en fréquence réduite. En plus, les orientations les plus importants sont ceux de 0 ° et 90 °, on prendra, donc, des orientations paires : 2, 4 et 6. Pour les fréquences, on va de 3 à 7. Comme on part de la fréquence maximale et, en général, les informations sont situées à basse fréquence, l'idée c'est de observer les résultats avec toutes les combinaisons, pour qu'on puisse prendre une décision sur quel combinaison est la plus performante. La figure 11 montre le résultat où une LDA a été faite sur le *dataset training* et les bases de *Lille* et *Grenoble* ont été classifiées dessus.

Les tableaux montrent le nombre de mauvaise classification. Comme ses deux *datasets* ont 200 images, ce serait le nombre maximum d'erreur. On voit que pour ce de *Grenoble*, figure 11a, 4 orientations c'est le cas idéal pour le classement et que, pour ce de *Lille*, figure 11b, la performance est plus distribuées entre les combinaisons et aussi le taux d'erreur c'est plus significatif, comme on attendait.

GrenobleBW_exp				
f0_max		orientations		
0,3		4	6	8
frequences	3	1	47	61
	4	0	34	53
	5	0	13	49
	6	0	2	20
	7	0	0	5
MIN:		0		
ERREUR MOYEN (%):		10%		
ERREUR MIN (%):		0%		

LilleBW_exp				
f0_max		orientations		
0,3		4	6	8
frequences	3	46	44	47
	4	48	47	51
	5	50	46	46
	6	51	45	47
	7	49	49	49
MIN:		44		
ERREUR MOYEN (%):		24%		
ERREUR MIN (%):		22%		

(a) Test du *dataset* de Grenoble

(b) Test du *dataset* de Lille

Figure 11: Tests de Grenoble et Lille sur le *dataset training*

On n'avait pas encore pris une décision sur la meilleure combinaison, d'une partie parce que les deux *dataset* ont de comportement très différents et d'autre qu'on voudrait travailler avec au moins 6 orientations, pour qu'on puisse, par exemple, dans un second temps, classer les images naturelles entre plages, montagnes, etc. Donc, on décide de garder toutes les combinaisons et on test un autre particularité qui c'est l'influence des différents normalisations sur nos résultats. Le tableau 12 montrent tous les essais qu'on a fait.

Dans ces *batches*, on a utilisé le *dataset training* comme apprentissage et aussi comme test. Pour regarder sur l'ensemble des images comme ce classifieur performe. La figure 13 montre l'exemple du cinquième *batch*, ce qui a été le plus performant en terme d'erreur, et aussi le sixième, celle le moins performant.

Batch	Norm 2	Norm Freq	Norm 1
1	✓	✓	✓
2	✗	✓	✓
3	✓	✗	✓
4	✗	✗	✓
5	✗	✓	✗
6	✗	✗	✗

Figure 12: *Batches* générés pour tester l’influence des normalisations

training					training				
f0_max		orientations			f0_max		orientations		
0,3		4	6	8	0,3		4	6	8
frequences	3	239	276	293	frequences	3	287	394	452
	4	240	296	312		4	305	391	438
	5	248	295	328		5	328	377	434
	6	256	303	332		6	342	356	416
	7	273	304	323		7	337	348	365
MIN:		239			MIN:		287		
ERREUR MOYEN (%):		18%			ERREUR MIN (%):		18%		
ERREUR MIN (%):		15%			ERREUR MOYEN (%):		23%		

(a) Test du *dataset training* - batch 5

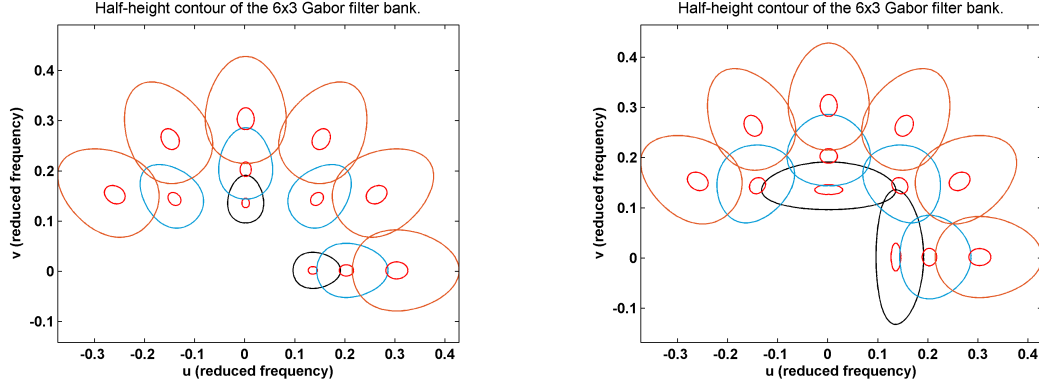
(b) Test du *dataset training* - batch 6

Figure 13: Tests de Grenoble et Lille sur le *dataset training*

7.1 Analyse des paramètres du banc de filtres de Gabor

Après ce premier test, on voit que les résultats le plus performant sont sans les 2 normalisations (étape 3 et 6 du modèle). On garde encore toutes les orientations et fréquences et on joue sur 2 paramètres du banc de filtre de Gabor :

1. On a testé ce type de filtre en ayant le nombre d’orientations par fréquence variable. Les deux images de la figure 6 montrent qu’il est constant. Pour le faire, on peut soit garder la bande transversale comme elle est, c’est-à-dire, elle considère la fréquence réduite actuel et le nombre maximum d’orientations, soit chaque couche prends son nombre d’orientations pour le calcul. La figure 14 montrent les deux démarche mis en oeuvre.
2. On a changé aussi sa bande radiale. Pour le moment, elle est constante et égale à 1, en octave. Néanmoins, les données psychophysiques montrent qu’elle varie en fonction de la fréquence spatiale [7]. La figure 15 montre cette variation.



(a) Banc de filtre de Gabor 6x3, avec la bande transversale calculée à partir du nombre maximum d'orientations. (b) Banc de filtre de Gabor 6x3, avec la bande transversale calculée à partir du nombre d'orientations de chaque couche.

Figure 14: Altérations sur le banc de filtre de Gabor

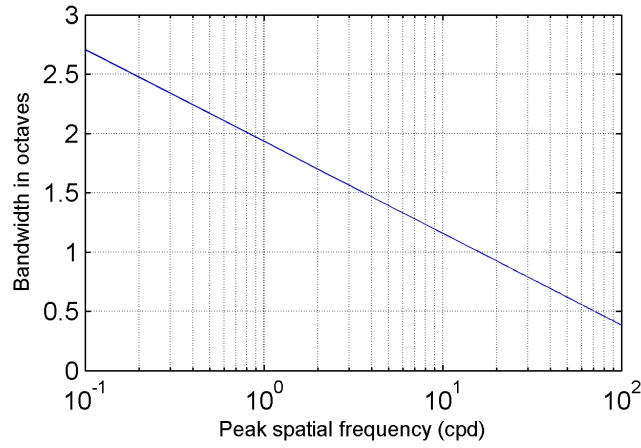


Figure 15: Variation de la bande radiale en fonction de la fréquence spatiale

7.1.1 Largeur de bande radiale

La figure 16 montre qu'on perd un peu en performance (augmentation de 18 % à 22 % de l'erreur moyen), mais ce modèle est plus réaliste et on décide de le garder par la suite.

training				
f0_max		orientations		
0.3		4	6	8
frequences	3	295	387	403
	4	286	371	411
	5	286	367	394
	6	283	356	396
	7	288	350	373
MIN:		283		
ERREUR MOYEN (%):		22%		
ERREUR MIN (%):		18%		

(a) Performance du classifieur avec la bande radiale variable

training				
f0_max		orientations		
0.3		4	6	8
frequences	3	239	276	293
	4	240	296	312
	5	248	295	328
	6	256	303	332
	7	273	304	323
MIN:		239		
ERREUR MOYEN (%):		18%		
ERREUR MIN (%):		15%		

(b) Meilleure performance jusqu'à ce moment.

Figure 16: Comparaison entre les performance du banc de filtre de Gabor avec la bande radiale constante et variable

7.1.2 Largeur de bande transversale

On a travaillé sur 2 possibilités de changement de la bande radiale, comme montré par la figure 14. Pour construire le banc de filtre de Gabor avec un nombre d'orientations variable par fréquence, le principe sera le suivant :

- 3 fréquences :
 - La première couche aura toujours 2 orientations.
 - La deuxième couche aura toujours 4 orientations.
 - La troisième couche aura 6 orientations, si possible. Si non, elle aura le nombre maximum d'orientations de la combinaison analysée¹.
- 4 fréquences :
 - La première couche aura toujours 2 orientations.
 - La deuxième couche aura toujours 4 orientations.
 - La troisième couche aura 5 orientations, si possible.
 - La quatrième couche aura 6 orientations, si possible.

La figure 17 montre les résultat. Pour comparer l'efficacité du changement, j'ai mis dans la figure 17a le performance du banc de filtre en question et dans la figure 17b les meilleures jusqu'à ce moment là. Deux dernières remarques, la bande transversale est calculée à partir du nombre maximum d'orientations, comme exemplifié par la figure 14a et la bande radiale est variable.

1. Dans le cas du filtre 4x3, le nombre maximum d'orientations est 4, donc la dernière couche l'aura. Dans le cas 8x3, la dernière couche aura 6 orientations, parce que 6 est inférieur au nombre maximum d'orientations.

training				
f0_max		orientations		
0.3		4	6	8
frequences	3	268	243	238
	4	264	256	240
MIN:		238		
ERREUR MOYEN (%):		16%		
ERREUR MIN (%):		15%		

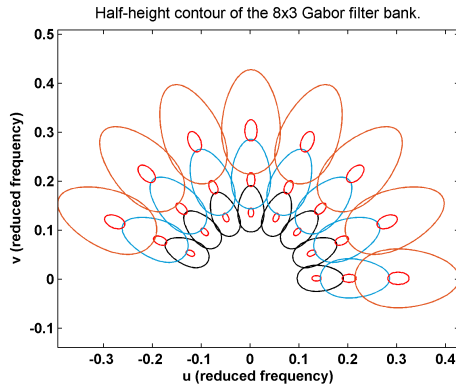
(a) Performance avec un nombre d'orientations variable par fréquence.

training				
f0_max		orientations		
0.3		4	6	8
frequences	3	239	276	293
	4	240	296	312
MIN:		239		
ERREUR MOYEN (%):		17%		
ERREUR MIN (%):		15%		

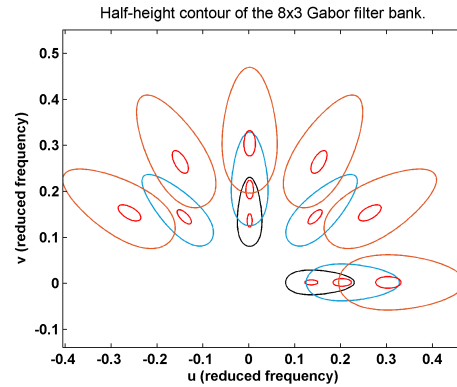
(b) Meilleure performance jusqu'à ce moment.

Figure 17: Comparaison entre les performances du banc de filtres de Gabor avec le même nombre d'orientations par fréquence et avec un nombre variable d'orientations par fréquence

Encore une fois, les tableaux montrent les nombres d'images mal classifiées parmi le total de 1600. La similarité des performances suggère qu'il y a des filtres qui n'ont pas une réponse significative pour la classification. La figure 18 donne une idée de la différence entre le nombres de banc des filtres utilisés pour 8 orientations et 3 fréquences.



(a) Filtre de Gabor 8x3, avec bande radiale variable et le nombre d'orientation par fréquence constant.



(b) Filtre de Gabor 8x3, avec bande radiale variable et le nombre d'orientation par fréquence variable.

Figure 18: Différences de quantité de banc de filtre de Gabor.

7.2 Approche “*imagette*”

Une autre démarche pour classer une image donnée, c'est de, à la place d'utiliser la chaîne de traitement dans l'image entière, couper l'image en plusieurs *imagettes* et on classifie chacune de ces pièces séparément. La figure 19 montre comment les pièces ont été choisi. L'idée c'est que la classification donnée par l'imagette centrale (1) aurait le poids le plus important et les autres 4 auraient un poids similaire. Ce vient du fait qui le modèle assume qui le regard est fixé sur le centre de l'écran et les traitements appliquées sur l'image prennent cette considération.

Cependant, comme la figure 20 montre, quand le même poids a été attribué pour toutes les imagerie, nos résultats ont été mieux.

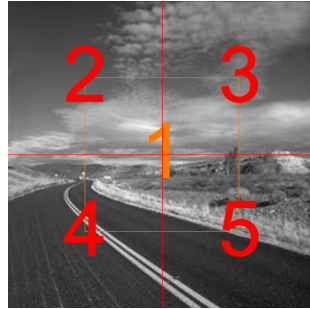


Figure 19: Division de l'image originale en 5 imagerie

training				
batch 15		Orientations		
		4	6	8
Frequencies	3	235	236	249
	4	236	243	260
	5	252	254	271
	6	266	259	279
	7	279	262	288
Min Error		235		
Avg Error (%)		16%		
Min Error (%)		15%		

(a) Performances avec la même probabilité pour toutes les imagerie.

training				
batch 15		Orientations		
		4	6	8
Frequencies	3	247	283	308
	4	274	296	319
	5	283	302	327
	6	289	307	337
	7	301	311	335
Min Error		247		
Avg Error (%)		19%		
Min Error (%)		15%		

(b) Performances avec une probabilité 4 fois plus élevée pour l'imagerie centrale.

Figure 20: Comparaison des performances avec différentes poids pour les imagerie.

7.3 Classification

Maintenant, on essaie tester le classifieur d'une manière plus systématique. Jusqu'à ce moment, il était utilisé juste pour comparer les résultats dus aux différents modification dans le modèle. Néanmoins, on veut vérifier sa performance en utilisant la méthode de validation croisée (*Cross Validation*).

7.3.1 Validation Croisée

Après avoir faite tout cet ensemble de tests, on est arrivé aux conclusions suivants :

1. On garde la largeur relative de la bande radiale variante, ce qui amène le modèle à une version plus réaliste.
2. On travaille encore sur un nombre d'orientations par bande de fréquence constant, comme la figure 18a montre.
3. On garde la sortie parvocellulaire, ce qui donne les résultats les plus performant et qui est aussi cohérent par rapport aux modèle de traitement du cerveau.

4. L'approche imagerie est une bonne alternative, puisque les résultats sont bons et on peut jouer sur les poids des images pour éventuelles améliorations des résultats.

On change, donc, notre méthode d'apprentissage et test. Au lieu d'apprendre et tester sur la même base d'image, ce qui était juste pour valider les étapes du modèle, on repart nos images en 2 parties. La partie apprentissage contiendra 2/3 des images et la partie test le reste. Ainsi, des nouvelles images seront tester et on peut avoir une idée plus réaliste des performances du classifieur.

À partir des conclusion mentionné précédemment, on prend aléatoirement la partie apprentissage et tests, en gardant la proportion de 50 % de chaque type, c'est à dire, 1/6 des images de tests seront naturelles et 1/6 seront artificielles. On fait 100 itérations différents et on est intéressé par la moyenne de l'erreur moyen sur toutes les itérations bien comme son écart-type. On tests deux options :

- Approche Globale : l'image entière est traitée.
- Approche Locale : l'image est réparti en 5 imagerie et, pour le moment, les poids des images seront égaux.

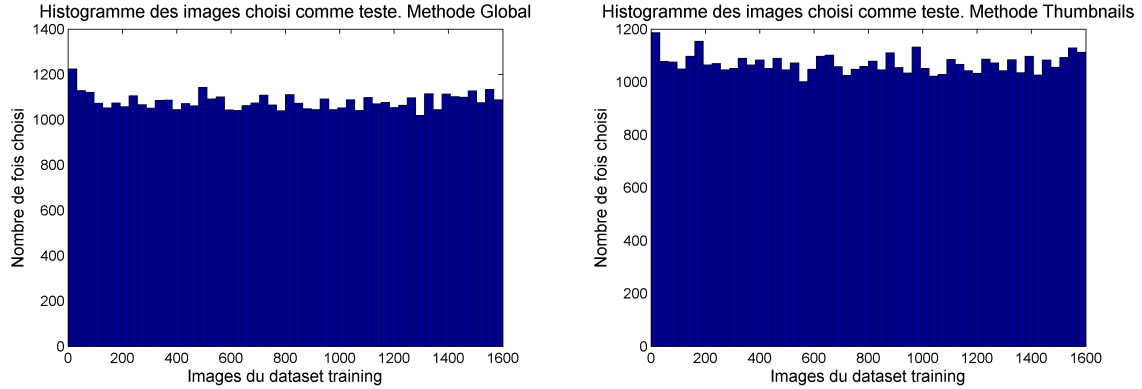
La figure 21 montre les résultats obtenus. On aperçoit que le nombre de mauvaise classification c'est similaire à ceux d'avant, mais, maintenant, on test 534 images, non plus toutes les 1600. Le tableau montre, pour chaque orientation x fréquence, la moyenne de mauvaise classification pour toutes les itérations. En plus, pour chaque itération on garde l'erreur moyen et, en regardant l'écart type de cet erreur, on voit qu'il est bien fixé sur 21 % pour les deux approches. La différence d'approche n'a pas apporté une différence en performance.

training - Approche Globale					training - Approche Locale				
batch 15		Orientations			batch 15		Orientations		
		4	6	8			4	6	8
Frequencies	3	114.255	115.81	115.88	Frequencies	3	115.89	112.755	113.225
	4	113.185	115.9	114.685		4	115.755	112.775	113.015
	5	112.465	114.715	115.805		5	115.33	112.375	112.52
	6	113.09	114.395	115.98		6	115.58	112.82	112.82
	7	113.08	114.375	115.33		7	114.795	112.6	114.585
Min Error		112.465			Min Error		112.375		
Avg of Avg Error (%)		21.4600%			Avg of Avg Error (%)		21.3089%		
Avg of Min Error (%)		20.7041%			Avg of Min Error (%)		20.7004%		
Ecart Type of Avg Err (%)		1.2738%			Ecart Type (%)		1.0799%		

(a) Performances du approche globale avec 100 itérations. (b) Performances du approche locale avec 100 itérations.

Figure 21: Comparaison des performances entres les méthodes globale et locale.

Une dernière remarque, la figure 22 montre l'histogramme des images choisit pour toutes les itérations. On regarde qui le choix des images n'a pas biaisé les résultats, dans le sens qui a été bien reparti entres toutes les 1600 images.



(a) Histogramme des images choisies pour l'approche globale. (b) Histogramme des images choisies pour l'approche locale.

Figure 22: Distribution des images choisies pour les tests.

7.3.2 Réseau Neurone

Une dernière essai c'était le changement de classifieur vers un réseau neurone. Pour le faire, on a profité de la générosité des messieurs Roman PIEKUT et Martial MERMILOD qui ont disponibilisé leur réseau neurone pour qu'on puisse vérifier les performances.

Dans cette partie, on change le nombre de catégorie par 8. C'est nécessaire pour qu'on s'assure que les résultats du réseau neurone sont fiables et non pas au hasard ou même due à une surapprentissage. Ainsi, la table 2 montre les nouvelles catégories créées. Chaque catégorie possède 200 images.

Table 2: Nouvelles catégorie créées.

Catégorie	Sous-catégorie
Artificielle	Autoroute
	Bâtiment
	Autoroute + bâtiment
	Immeubles
Naturelle	Plage
	Forêt
	Montagne
	Plaine

Pour utiliser le réseau, on a choisi le banc de filtre de Gabor 8 orientations et 7 fréquences. En plus, il possède 8 couches de entrée (l'énergie du banc de filtre de Gabor pour une fréquence donnée), 8 couches cachées et 8 couches de sorties (8 catégories possibles). Le réseau peut être réglé de 2 manières différents : soit l'information est traitée progressivement, c'est-à-dire, on n'a pas de réinjection d'information, ce qui

s'appelle Multilayer Perceptron (MLP). La deuxième option c'est le Simple Recurrent Network (SRN), qui permet un comportement dynamique temporelle [8].

Dans le démarche SRN, on a testé 2 types d'interaction : Le Fine-to-Coarse (FtC) présente les fréquences les plus basses au réseau et après les plus hautes et le Coarse-to-Fine (CtF) fait le processus inverse. Comme on sait que dans le cerveau on reçoit l'information basse et haute fréquence dans des instants séparés [4], c'est bien de regarder comment le réseau réagit.

La figure 23 montre les résultats. Comme depuis le début, ce qui nous intéresse c'est la classification en naturelle ou artificielle, la pourcentage de bonne classification que la figure montre est liée à la catégorie classifiée, et ne pas à la sous-catégorie, comme la convention montrée dans la table 2. Si, par exemple, une Plaine est classifiée comme une Plage, ça compte comme une bonne classification parce que les deux sous-catégories appartiennent à la catégorie Naturelle.

On voit que dans le trois démarches testées, le MLP possède le résultats plus performants, principalement à la fréquence réduite de 0,133. On est arrivé à environ 90 % de bonne classification, une chiffre qui est très remarquable étant donnée la simplicité du méthode qui prend en compte, dans ce cas, $8 \times 7 = 56$ valeurs d'énergie d'une image 255×255 pour accomplir la tâche de classification.

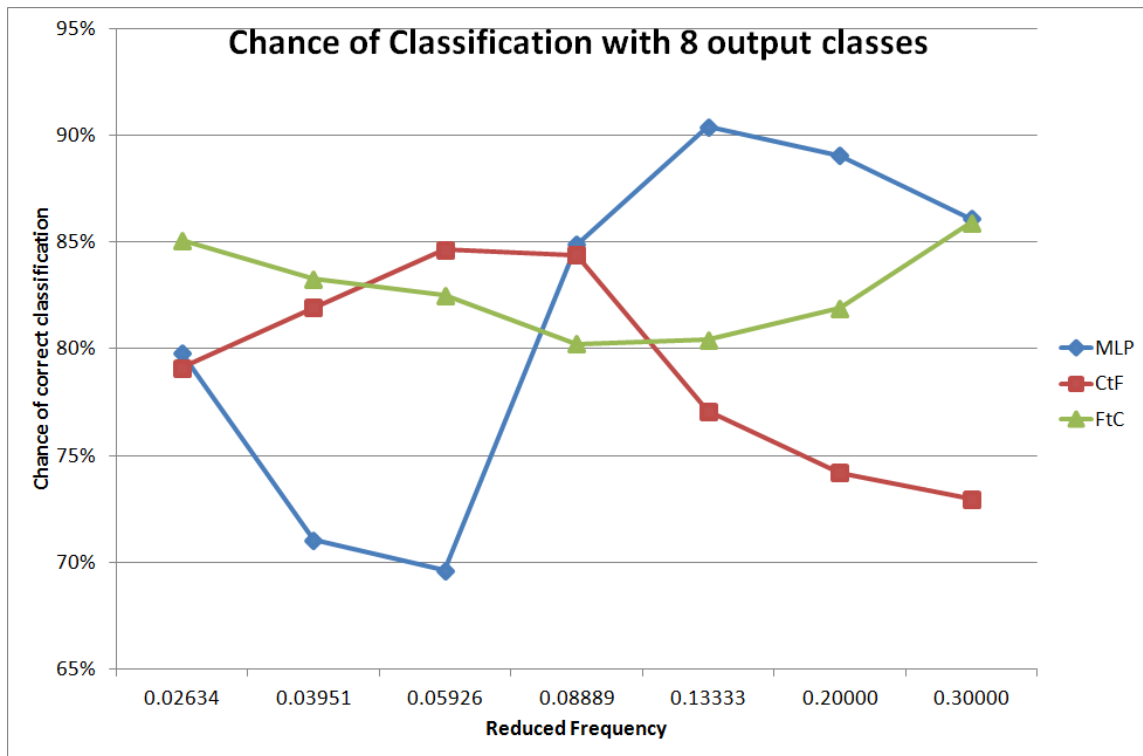


Figure 23: Résultats de bonne classification du réseau neurone.

8 Conclusion

L'expérience du stage a été très enrichissante, dès le jour où j'étais accepté pour le poste, avec un sujet très intéressant par rapport à ce qui je cherchais mais aussi un peu vague ou abstraites à cause de la manque des plusieurs connaissances pour pouvoir avoir une idée plus concrète du projet.

Du côté technique, il est vraiment lié à la filière, pas mal de fois j'ai pu retrouvé des contenus étudiés en cours, comme l'ACP et l'LDA, le *zero-padding*, la TF, etc. En plus, comme je savais mon choix pour le IESE 5 dans l'option Automatique, Traitement des Signaux et Images (ISA), c'était une bonne opportunité pour avoir un contact plus proche avec cette spécialisation.

Du côté personnel, l'équipe était très disponible, l'ambiance très conviviale et je espère qu'on garde le contact parce que les portes seront toujours ouvertes si ça ne dépend que de moi.

Du côté résultats, je crois qu'on a atteint ce qu'on s'est proposé dès le début du stage. J'ai pu comprendre le modèle et le modifier selon les nécessités. On a testé plusieurs cas de figures pour voir l'impact de chaque étape de traitement dans le modèle de perception visuelle. En plus, on a pu tester à quel point LDA est capable de faire une classification entre scènes naturelles et artificielles en faisant une validation croisée. Pour finir, on a obtenu des résultats très intéressants avec le réseau neurone où avec si peu d'information sur l'image, on a eu environ 90 % de bonne classification.

Pour l'avenir, on a encore plusieurs points à travailler dessus :

1. Mise en oeuvre d'une méthode d'optimisation pour trouver les poids les plus adaptés pour chaque image.
2. Utiliser la partie basse fréquence du banc de filtre de Gabor pour l'approche globale et la partie haute fréquence pour l'approche locale.
3. Utiliser le banc de filtre de Gabor complet pour l'approche globale et une version pondérée de lui, selon les énergie obtenues, pour l'approche locale.
4. Comprendre comment utiliser les images sur un réseau neurone.

Il faut dire que je suis bien satisfait avec le résultat du travail, le défi de faire une tâche que on ne maîtrise pas semble parfois difficile, mais quand on travaille avec les gens qui sont toujours prêt à vous aider, la conquête est vraiment gratifiant.

Acronymes

ACP Principal Component Analysis. 24

CtF Coarse-to-Fine. 23

FtC Fine-to-Coarse. 23

ISA Automatique, Traitement des Signaux et Images. 24

LDA Linear Discriminant Analysis. 14, 24

MLP Multilayer Perceptron. 22

SRN Simple Recurrent Network. 22, 23

SUN Scene UNderstanding. 5

SVM Support Vector Machine. 5, 12, 14

TF transformée de Fourier. 8, 10, 11, 14, 24

ViBS Vision and Brain Signal Processing. 2

Références

- [1] Aude Oliva and Antonio Torralba. Modeling the shape of the scene : A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3) :145–175, 2001.
- [2] Aude Oliva and Antonio Torralba. Scene-centered description from spatial envelope properties. In HeinrichH. Bülthoff, Christian Wallraven, Seong-Whan Lee, and TomasoA. Poggio, editors, *Biologically Motivated Computer Vision*, volume 2525 of *Lecture Notes in Computer Science*, pages 263–272. Springer Berlin Heidelberg, 2002.
- [3] Aude Oliva and Antonio Torralba. Chapter 2 building the gist of a scene : the role of global image features in recognition. In L.M. Martinez J.-M. Alonso S. Martinez-Conde, S.L. Macknik and P.U. Tse, editors, *Visual Perception Fundamentals of Awareness : Multi-Sensory Integration and High-Order Perception*, volume 155, Part B of *Progress in Brain Research*, pages 23 – 36. Elsevier, 2006.
- [4] N. Guyader. Scènes visuelles : Catégorisation basée sur des modèles de perception. pages 1–46, 2004.
- [5] Support Vector Machines. <http://scikit-learn.org/stable/modules/svm.html>. Accédé le 29 mai 2017.
- [6] Linear Discriminant Analysis. http://sebastianraschka.com/Articles/2014_python_lda.html. Accédé le 08 juin 2017.
- [7] T. Ho-Phuoc. Developpement et mise en oeuvre de modèles d’attention visuelle. pages 44–45, 2010.
- [8] Tariq Rashid. *Make your own neural network*. CreateSpace Independent Publishing Platform, 2016.

Étudiant : LEMOS Lucas

Année d'étude dans la spécialité :
4ème

Entreprise : GIPSA-lab

Adresse complète : 11 rue des Mathématiques

Grenoble Campus BP46

F - 38402 Saint Martin d'Hères Cedex

Téléphone : +33 04 76 82 71 31

Responsable administratif : Lucia BOUFFARD-TOCAT

Téléphone : +33 04 76 57 43 50

Courriel : rh@gipsa-lab.grenoble-inp.fr

Tuteur de stage : GUYADER Nathalie

Téléphone : +33 04 76 57 43 72

Courriel : nathalie.guyader@gipsa-lab.grenoble-inp.fr

Enseignant-référent : LAVAUT Yves

Téléphone : +33 06 03 45 13 03

Courriel : yves.lavault@aii-biomedical.com

Titre : Modèle bio-inspiré de perception visuelle pour la catégorisation d'images
de scènes naturelles

Résumé : La vision est une fonction primordiale pour l'être humain. Ce projet propose une validation d'un modèle existant d'un système de perception visuelle chez l'homme. Cette système a été développé sur Matlab [4] et l'objectif principal c'est de quantifier l'impact de chaque étape de la chaîne de traitement du modèle par rapport au pourcentage de bonne classification des images en 2 catégorie : naturelle et artificielle. On commence par une explication détaillé de chaque étape, on présente les images qui seront utilisé bien comme le classifieur choisi. Après, on test les étapes et différents proposition pour l'amélioration des résultats et qui amène le modèle de perception visuelle plus proche de la réalité. On montre les donnés obtenus avec une validation croisée et, pour finir, on utilise un classifieur plus puissant pour voir jusqu'où on est capable de bien classifier les images étant donné la petite quantité d'information prise en compte par le modèle.