

Signaux Images et Modèles de Perception Visuelle

M2 CNA – 2021-2022

Louise Kauffmann

CM1

Informations pratiques

- **Contact**

Louise Kauffmann (MCF UGA, Laboratoire de Psychologie et Neurocognition)

louise.kauffmann@univ-grenoble-alpes.fr

Bureau E105, Bâtiment Michel Dubois

- **Organisation des cours**

12h de CM et 6h de TP

3CM assurés par LK → Perception visuelle de la rétine au cortex, outils de traitement de signal

1CM + 1TP assurés par Nathalie Guyader → modélisation des traitements rétiniens et corticaux

2CM + 1TP assurés par David Alleyson (CR CNRS, LPNC) → Modélisation et perception de la couleur

- **Examen**

20%CC → Compte rendus TP

80%ET → Examen de 2h avec questions de réflexion sur le cours et sur des résultats d'expériences

Contexte général du cours

- Cours créé en ~2010 par Nathalie Guyader et Alan Chauvin
- **Objectifs: A partir des connaissances sur le fonctionnement du système visuel, quels modèles pour simuler la perception visuelle?**
 - La perception visuelle des primates est un modèle efficace, robuste, spécialiste, générique et adaptif
 - C'est encore le modèle le plus complet pour la « reconnaissance visuelle »

- **A quoi cela sert d'étudier la perception visuelle ?**

- Meilleure compréhension du système visuel
 - *Applications cliniques (aide au diagnostic etc.)*
- Suppléance sensorielle / comportementale
 - *Prétraitements visuels*
- Amélioration de la vision par ordinateur
 - *Surveillance, reconnaissance d'action*
 - *Compression (images, vidéos)*
 - *Codage efficace de la couleur*

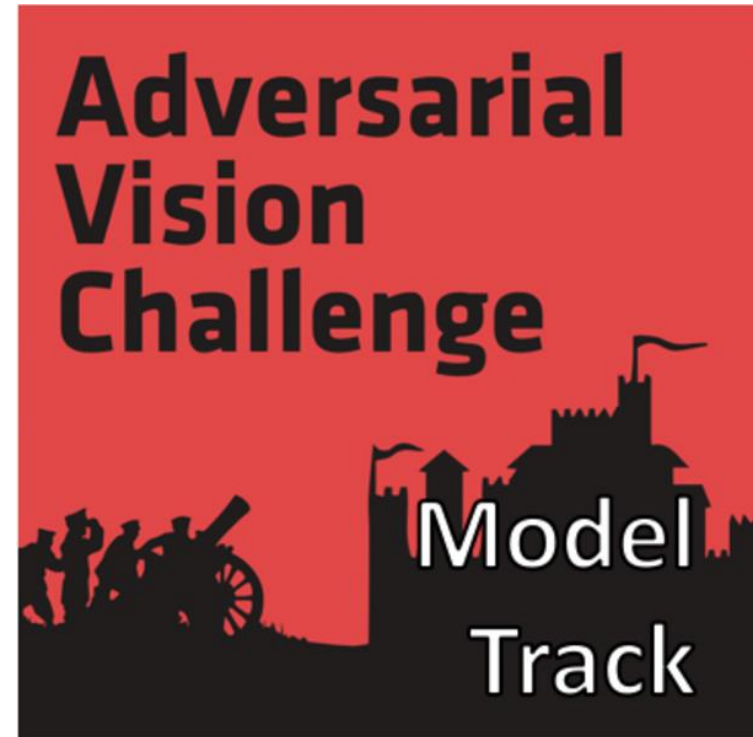


Un exemple...

NIPS 2018 (*Neural Information Processing Systems*)

Welcome to the *Adversarial Vision Challenge*, one of the official challenges in the [NIPS 2018](https://www.crowdai.org/challenges) competition track. In this competition you can take on the role of an attacker or a defender (or both).

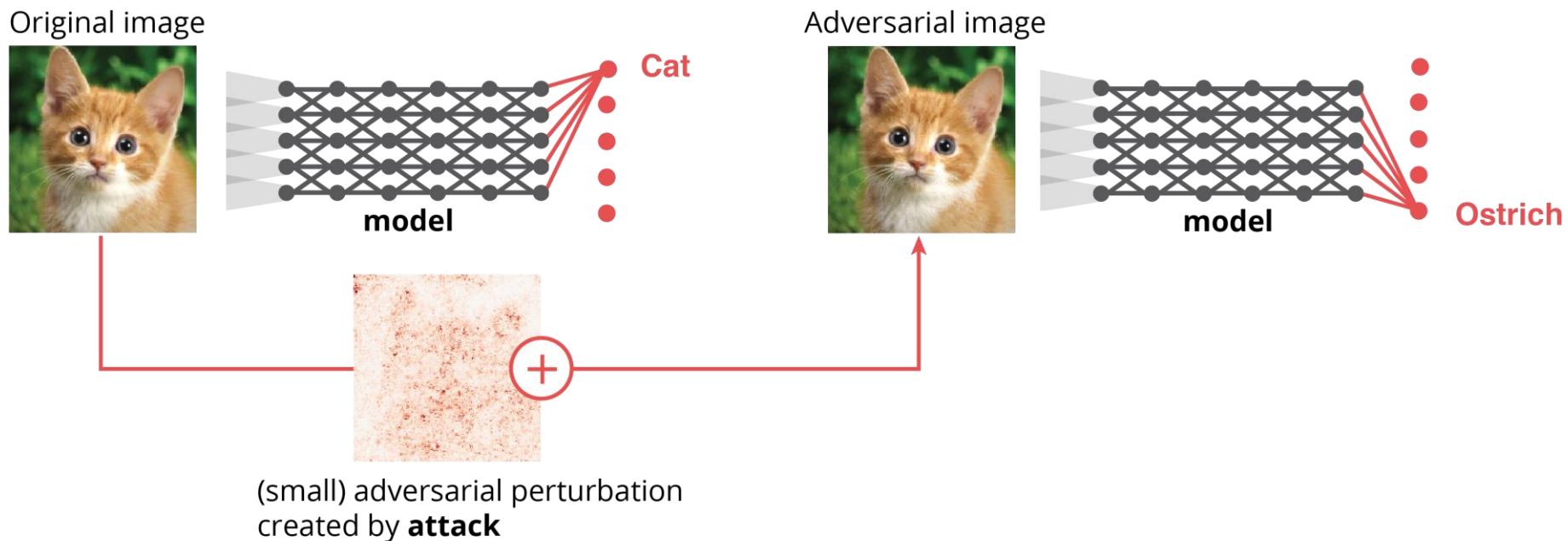
<https://www.crowdai.org/challenges>



Un exemple...

NIPS 2018 (*Neural Information Processing Systems*)

- As a **defender** you are trying to build a visual object classifier that is as robust to image perturbations as possible.
- As an **attacker**, your task is to find the smallest possible image perturbations that will fool a classifier.



Comment étudier la perception visuelle?

- **Mesures comportementales et physiologiques**

- Neurosciences : mise en relation d'une stimulation et d'une réponse d'un neurone ou d'un groupe de neurones (imagerie et neuro-stimulation)
- Psychologie/Psychophysique : mise en relation d'une stimulation et d'une réponse comportementale (Protocole expérimental, plan d'expériences et variables)

- **Modélisation**

- Analyse des propriétés de l'environnement : traitement du signal, théorie de l'information
- Modélisation (conceptuelle ou informatique) pour proposer des prédictions non observées: « computer vision », modélisation bio-inspirée (fonctionnement des cellules du système visuel) etc.

- **Quelle est la stimulation idéale pour étudier le système visuel ?**

- Des stimulations dont on contrôle toutes les dimensions : points, sinusoïdes, etc.
- Des stimulations écologiques c'est-à-dire des stimulations auxquelles le système visuel s'est adapté : le monde réel

Comment étudier la perception visuelle?

- **Ce qui ne sera pas traité:**
 - Aspects physiques de la perception visuelle (optique, relations surface-illuminants...)
- **Ce qui sera traité:**
 - Étapes de traitement de l'information visuelle pour
 - Chercher un modèle de scène
 - Chercher un modèle de perception de scènes

La perception visuelle

Définition

Processus de transformation d'un signal lumineux en un objet interprétable par le système neuronal, moteur, mnésique ou sémantique

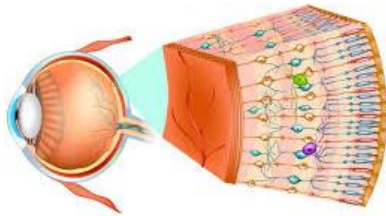
Un percept visuel est la première étape de la reconnaissance (consciente ou non)

Etapes :

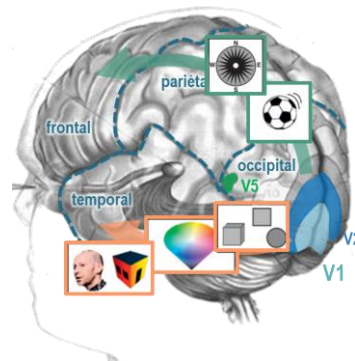
Scène visuelle



→ réception



→ extraction/analyse



→ interprétation

« Plage »

« Vacances »

Qu'est ce qu'une scène visuelle?

Exemple: Comment trouver rapidement une image parmi des milliers d'autres?

- S'inspirer des systèmes utilisés pour gérer les textes: Chaque texte est décrit par un nombre réduit de mots clés en relation avec son contenu et son style. Cette réduction d'information permet de retrouver plus rapidement un texte ou un ensemble de textes ayant des « traits » communs. Ce principe est aujourd'hui utilisé par les moteurs de recherche sur le web.
- Appliqué à l'image, cela consiste à extraire une information réduite de l'image tout en conservant une représentation la plus complète possible. **Quelle est cette information?**

Qu'est ce qu'une scène visuelle?

Vision naïve

Google image scene naturelle

Tous Images Vidéos Actualités Shopping Plus Paramètres Outils Collections SafeSearch

istock couleurs vives peinture 123rf dreamstime vecteur verdure printemps depositphotos forest paysage saisonnière accroche arbre oiseaux tropicale

Belle Scène Naturelle De La Forêt De V...
fr.123rf.com

Naturel, inondation, désastre, ...
canstockphoto.fr

Parc De Ville. Paysage D...
fr.123rf.com

Scène Naturelle De L'eau De Style, P...
fr.dreamstime.com

Scène naturelle de la for...
fr.depositphotos.com

Scène Naturelle Avec La Silhouette De L'...
fr.123rf.com

Photo libre de droit de La Forêt Nature P...
istockphoto.com

Belle Scène Naturelle Des Ar...
fr.pngtree.com

Scène Naturelle De La Forêt...
istockphoto.com

Photographie Couleur Détail Allée Ch...
fr.depositphotos.com

Scène Naturelle Vecteur Scène. Montagne...
fr.123rf.com

Painting by Numbers Printemps Scèn...
amazon.fr

Scène Naturelle Entraînante De Sud...
fr.dreamstime.com

Naturel Promotionnel Scène...
french.alibaba.com

Que voyez-vous? Comment décririez-vous ces scènes?

→ *Tendance à « labelliser » les scènes, les objets qui les composent*

Qu'est ce qu'une scène visuelle?

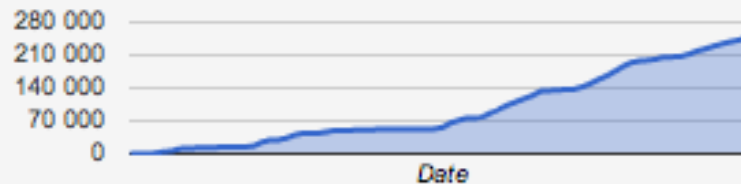
Vision collaborative

- Site web: <http://groups.csail.mit.edu/vision/SUN/>
- Une grande base d'images
- Des participants

SUN Database: Large-scale Scene Recognition from Abbey to Zoo

Jianxiong Xiao James Hays[†] Krista A. Ehinger Aude Oliva Antonio Torralba
jxiao@csail.mit.edu hays@cs.brown.edu kehinger@mit.edu oliva@mit.edu torralba@csail.mit.edu
Massachusetts Institute of Technology [†]Brown University

SUN
database



131072 Images
908 Scene categories

Qu'est ce qu'une scène visuelle?

Vision collaborative

Indoor **Urban** **Nature**

airlock armoury access road alleyway aqueduct arbor archeology

anechoic chamber

Catégories génériques ou catégories de base

bookbindery	bowling	departure lounge	campus	carport	fly bridge	crag	cromlech	ditch
dais	boat deck house	jewelleryshop	fire escape	floating bridge	lookout station	glen	gorge	grassland
hatchway	hunting lodge	launchpad	loading dock	porch	marsh	mineshaft	mountain	

parlor	pilohouse	police office	piazza	plantation	porch	rice paddy	river	rock outcrop
skating rink	sports stadium	staircase	shelter	signal box	skyscraper	snowbank	stream	sunken garden

Catégories spécifiques ou subordonnées

Qu'est ce qu'une scène visuelle?

Vision collaborative

Limite:



« village »



« mer »



« plage »

Qu'est ce qu'une scène visuelle?

Vision collaborative

Limite:



Régression → Une scène est une collection de scènes...

Qu'est ce qu'une scène visuelle?

Vision collaborative

Limite:



Régression → Une scène est une collection d'objets...

Qu'est ce qu'une scène visuelle?

Vision collaborative

- Autre outil collaboratif lui aussi développé au MIT:



Welcome to LabelMe, the open annotation tool.

The goal of LabelMe is to provide an online annotation tool to build image databases for computer vision research. You can contribute to the database by visiting the annotation tool.

Log In

Username

Password

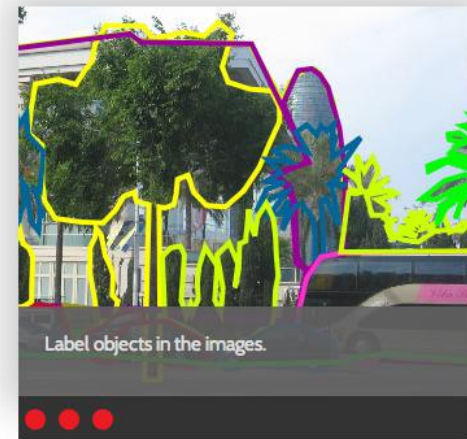
[Forgot your password?](#)

or Sign Up [Why?](#)

Name

Institution

..



<http://labelme.csail.mit.edu/Release3.0/>

LabelMe: Online Image Annotation and Applications

By developing a publicly available tool that allows users to use the Internet to quickly and easily annotate images, the authors were able to collect many detailed image descriptions.

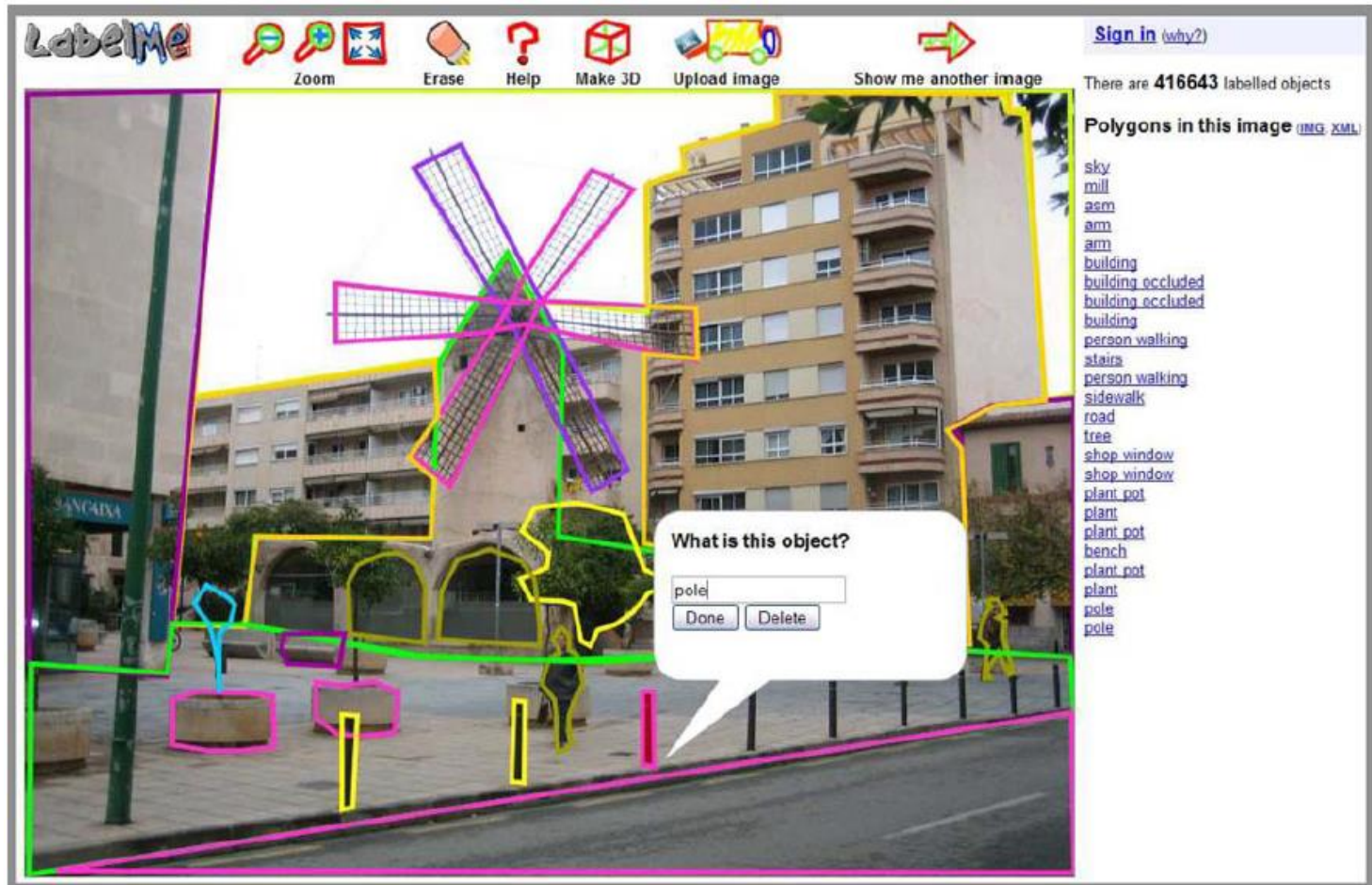
By ANTONIO TORRALBA, BRYAN C. RUSSELL, AND JENNY YUEN

LabelMe: a database and web-based tool for image annotation. B. Russell, A. Torralba, K. Murphy, W. T. Freeman. International Journal of Computer Vision, 2007.

Qu'est ce qu'une scène visuelle?

Vision collaborative

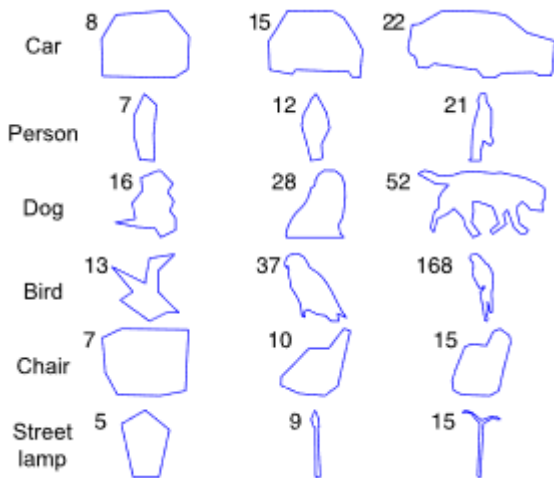
- Autre outil collaboratif lui aussi développé au MIT:



Qu'est ce qu'une scène visuelle?

Vision collaborative

- ➔ Approche descriptive et structurale dans le contexte de ces outils collaboratifs
- ➔ Une scène = une image, décrite par sa catégorie, sa segmentation structurale et ses labels



Qu'est ce qu'une scène visuelle?

Vision collaborative

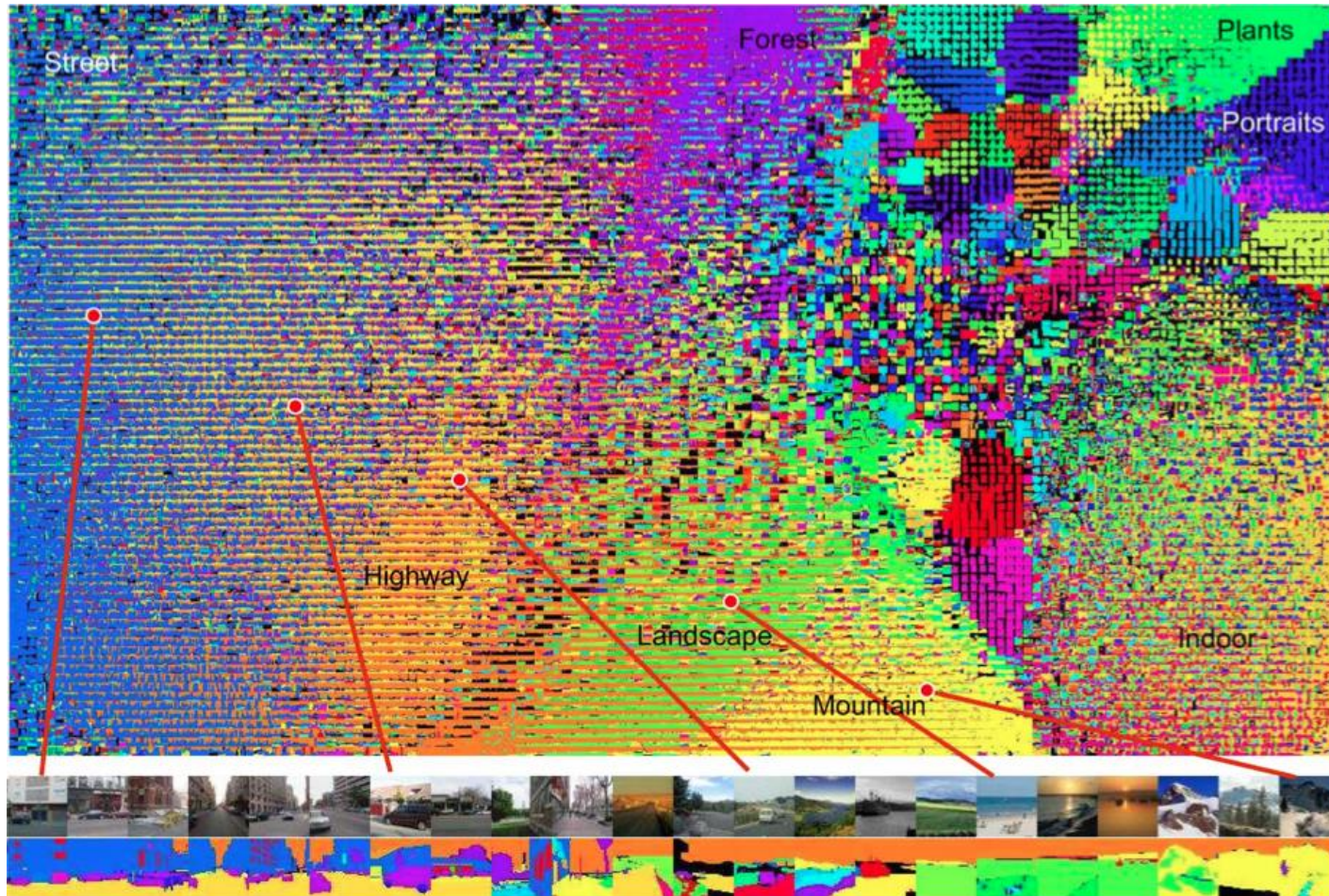


Fig. 6. *The images are arranged according to semantic similarity between images (nearby images will contain similar objects in similar spatial configurations). Each thumbnail shows the object segments of each image, with the objects consistently colored across the database. Although there are some easily identifiable clusters in the space, most of the images are organized across a continuous space in which transitions across images are smooth.*

Allons plus loin...

Une Plage



- Une scène

Allons plus loin...

Une Plage



- Une scène
- Une collection d'objets

Allons plus loin...

Une Plage



- Une scène
- Une collection d'objets
- Une collection de collection d'objets
- ...

Allons plus loin...

Une Plage



- Une scène
- Une collection d'objets
- Une collection de collection d'objets
- Des fonctions/ interactions potentielles avec l'observateur (notion « d'affordance »)

→ Limite d'une approche descendante

Synthèse: Une scène visuelle

- Pour éviter le problème de régression et de définition de l'objet, on regarde s'il existe des similitudes entre des régions classifiées comme similaires.
Plutôt que de définir à priori les régions comme des objets, on regarde si des régions qui se ressemblent appartiennent à la même chose
- Ainsi, au lieu de chercher une annotation dans l'objet on va la chercher dans les propriétés visuelles des scènes
Les features, traits ou « code book » visuel

Une scène = une collection de propriétés?

Une Plage



- Une scène
- Une collection d'objets
- Une collection de collection d'objets
- Des fonctions/ interactions potentielles avec l'observateur (notion « d'affordance »)
- **Des propriétés ou « traits visuels » (« features »)**

Dans ce cours...

Méthode: adopter une approche ascendante

→ Aboutir à une description d'une scène naturelle (reconnaissance d'objets) pour permettre de la catégoriser ou de prédire ses régions d'intérêts en s'inspirant du système visuel

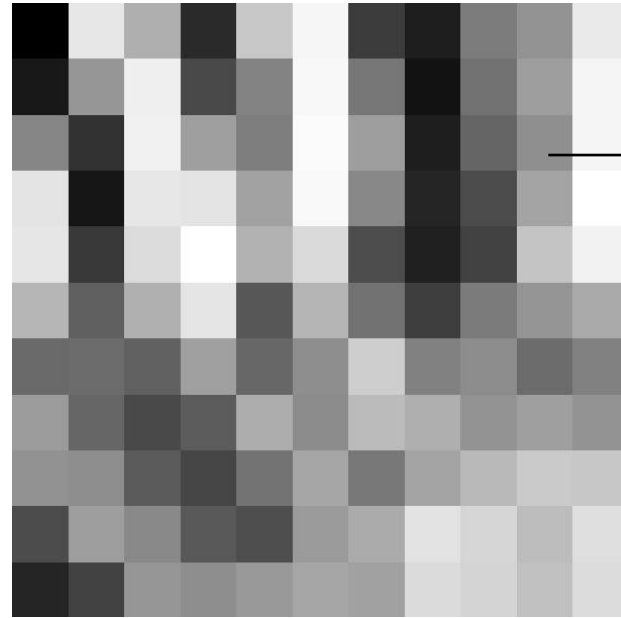
Outils: traitement d'images

- Considérer la rétine comme un échantillonneur discret de l'environnement
- Considérer qu'une image numérique est un modèle de l'image rétinienne
- S'intéresser aux traitements rétiniens et corticaux pour aboutir à une description de l'image

Qu'est-ce qu'une image?

46	47	47	47	46	45	44	44	40	48	53	44	106	106	59	72	126	129	135	83	77	131	115	73	90	155	169	166	169	145	143	171	175	175	163	157	138	137	145	150	146	146	180	182	171	167	165	158	161	166	167	171	173	167	169	184	180
48	49	50	49	47	45	45	44	48	53	56	48	112	113	66	76	138	128	125	76	68	116	120	115	125	164	163	158	158	141	145	164	169	168	160	161	147	145	145	142	155	147	173	176	173	174	173	164	156	162	164	166	165	156	157	171	173
49	51	53	52	50	47	47	47	52	55	45	49	113	111	71	77	137	128	130	88	80	114	115	124	130	158	159	160	160	147	152	164	166	164	158	159	145	144	148	145	150	146	176	177	170	168	167	159	157	166	163	168	167	162	166	181	179
48	51	54	54	52	50	49	50	52	54	53	49	111	113	71	77	137	137	134	84	81	120	117	123	128	163	167	163	166	143	142	158	160	163	160	159	136	136	144	148	153	150	149	181	183	173	163	167	150	155	164	158	155	138	160	176	178
48	51	54	54	52	50	50	50	52	56	55	52	110	111	77	80	139	141	133	78	79	123	121	127	128	171	169	164	165	138	136	163	156	160	162	161	133	132	148	152	156	148	174	176	173	174	171	162	153	164	159	153	162	170	184	177	
50	51	53	53	52	50	49	49	51	56	53	50	108	106	76	78	134	137	134	84	82	120	117	126	121	168	163	158	166	144	141	170	163	161	159	158	131	133	149	150	152	147	176	177	169	167	166	158	146	158	160	156	153	162	166	175	171
51	51	51	51	52	51	50	48	49	54	52	51	102	99	71	69	136	135	129	78	76	119	120	127	122	166	167	163	165	145	142	161	167	162	159	158	130	133	150	148	152	153	186	185	170	164	164	161	160	170	167	158	153	157	158	159	
51	50	49	50	52	53	51	49	49	53	48	39	95	67	63	126	129	127	76	79	129	126	119	120	161	169	166	160	144	149	162	161	160	165	167	136	134	147	143	157	151	176	174	165	164	166	162	162	171	166	158	162	165	164	164	166	
53	51	51	52	52	51	50	52	53	56	50	48	101	103	64	71	134	127	131	78	70	126	129	120	117	167	165	167	164	147	148	165	158	156	165	167	137	134	146	145	155	162	143	151	159	153	158	145	134	149	144	146	141	139	140	148	149
52	51	52	52	50	48	48	51	52	55	51	50	102	104	65	72	136	131	135	79	73	133	132	118	115	165	162	163	161	145	147	165	166	161	161	161	134	136	147	143	154	150	153	155	155	158	149	151	153	140	143	138	134	137	149	148	
50	51	53	52	49	46	47	49	52	54	52	51	102	105	66	72	138	133	138	84	75	131	130	118	116	168	162	160	157	141	142	162	168	165	159	156	129	135	148	144	150	141	170	169	149	156	155	150	136	135	131	151	151	138	132	142	143
50	51	52	51	49	48	47	48	50	52	51	52	101	104	67	70	139	132	140	88	75	124	125	121	114	169	163	161	159	142	141	163	164	170	162	160	130	136	151	150	149	136	160	161	151	160	153	146	138	131	158	156	137	128	137	140	
51	51	51	51	52	52	50	48	49	49	51	52	99	102	66	68	140	134	140	85	75	130	130	119	106	167	164	163	165	148	146	169	162	172	162	162	135	145	157	154	152	137	132	139	152	161	154	143	138	133	138	163	165	145	135	140	141
52	51	51	51	54	55	53	49	48	47	50	52	96	100	66	66	139	138	139	78	76	142	138	112	104	169	168	167	169	150	145	168	161	169	151	162	142	148	153	146	147	147	133	138	152	150	147	144	133	131	139	167	163	150	143	143	142
51	52	52	52	54	55	53	50	49	47	50	52	94	98	65	65	139	135	137	79	76	139	134	109	104	173	172	169	171	148	140	162	160	165	140	154	134	138	142	135	139	148	142	140	140	132	136	144	149	141	134	145	139	142	148	148	147
50	53	54	53	53	54	53	51	50	47	51	52	93	98	66	65	138	129	134	84	76	125	121	110	102	172	171	163	170	147	138	161	159	163	137	150	126	127	133	132	138	134	128	122	119	121	131	142	144	139	132	141	138	150	158	154	156
48	53	48	48	50	49	51	50	50	49	55	52	105	102	61	67	134	132	132	85	83	107	107	106	102	160	158	161	175	150	145	160	153	151	154	156	126	125	136	131	136	121	120	129	128	126	119	126	142	138	147	138	162	173	164	160	
46	53	48	47	48	46	49	50	49	47	53	52	108	105	61	64	135	135	135	87	84	106	104	107	103	163	162	163	174	151	146	160	152	152	157	156	131	126	133	125	137	132	111	112	126	128	125	116	118	132	133	158	136	152	181	155	142
44	52	49	47	46	43	47	49	49	46	51	50	107	106	63	65	133	136	135	87	81	103	102	111	103	163	166	164	173	152	147	160	155	155	156	159	131	125	132	125	132	128	115	117	125	124	122	117	122	124	150	175	184	140	159		
42	52	51	48	45	41	45	48	51	48	51	47	104	105	66	70	131	136	132	83	74	98	97	109	104	162	169	165	171	153	148	161	159	157	154	152	125	122	131	129	125	132	137	137	127	115	119	129	131	133	126	109	110	146	145	138	166
42	53	52	50	47	42	45	47	50	50	55	49	102	103	66	73	135	133	132	82	70	98	98	107	108	163	172	167	169	152	147	160	159	158	152	149	124	128	128	130	137	147	143	125	112	116	130	138	134	126	110	108	131	132	138	156	
43	54	53	52	51	45	46	46	47	50	51	60	52	101	63	70	136	138	133	87	74	114	115	116	115	163	174	169	168	151	144	160	155	158	154	152	130	123	126	124	138	131	134	131	122	124	131	128	113	113	152	160	151	151	124	126	
44	53	52	54	54	49	48	45	47	53	61	52	100	98	61	68	129	130	128	87	75	126	123	120	118	159	171	168	167	149	142	159	154	155	155	154	134	128	130	129	137	124	126	124	121	127	129	131	128	124	126	163	156	128	142	129	109
45	53	51	54	56	51	48	44	50	54	60	49	97	97	60	67	123	122	121	82	70	126	128	111	116	154	167	165	165	147	140	160	157	160	157	150	151	133	131	137	140	134	126	134	130	121	125	125	128	147	159	153	158	144	134	156	158
45	51	53	51	51	49	46	47	48	48	52	51	98	102	62	66	113	104	114	85	76	134	131	105	113	166	161	163	165	145	142	163	163	162	152	152	129	133	143	141	131	118	131	134	125	122	119	129	145	164	161	155	141	118	118	164	161
45	51	52	50	50	48	46	47	40	44	47	52	102	104	66	67	89	84	92	71	66	116	117	105	106	156	156	148	146	119	125	129	146	138	128	136	130	124	136	132	124	136	134	122	136	137	155	158	143	155	158	143	142	165	164		
45	51	52	49	49	48	46	47	40	47	44	53	102	100	68	65	91	90	90	73	69	105	107	105	106	156	156	148	146	119	125	129	146	138	128	136	130	124	136	134	122	136	137	155	158	143	155	158	143	142	165	164	167	164	176	157	
44	50	51	49	50	48	45	46	46	46	46	45	51	97	93	68	67	88	93	88	74	76	108	111	104	106	156	156	148	146	119	125	129	146	138	128	136	130	124	136	134	122	136	137	155	158	143	155	158	143							

Qu'est-ce qu'une image?



$$I [i,j] = N$$

i indice de ligne

j indice de colonne

N = Intensité du pixel

=

31	223	176	66	197	235	81	56	134	153	225
51	156	229	92	140	237	130	46	126	162	234
142	73	231	163	136	239	162	56	115	150	234
220	50	223	220	166	238	144	63	94	167	243
222	79	214	243	179	212	95	58	86	194	232
182	111	177	221	104	181	126	83	133	155	172
119	121	112	164	117	149	202	138	148	121	138
161	116	92	108	175	147	186	176	153	163	153
152	149	107	89	127	169	131	167	185	199	196
94	162	146	105	96	160	173	219	209	188	216
63	86	156	149	158	169	165	213	207	191	214

Une image = **signal 2D discret**

Définie avec une certaine **résolution**
et une **quantification**

Pixel = « picture element » élément de base d'une image

Résolution



256 x 256 pixels



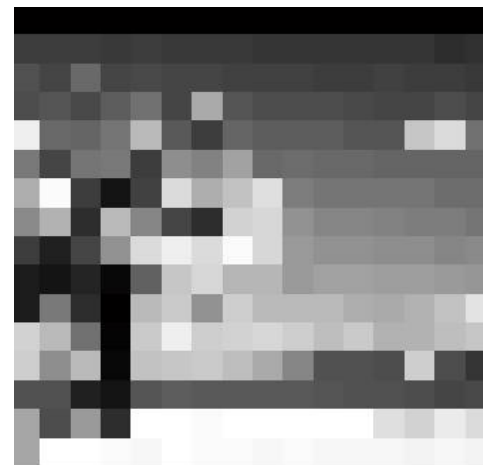
128 x 128 pixels



64 x 64 pixels



32 x 32 pixels



16 x 16 pixels

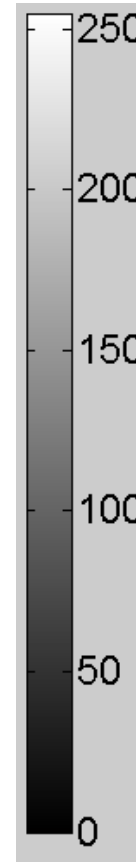
Quantification

Quantifier:

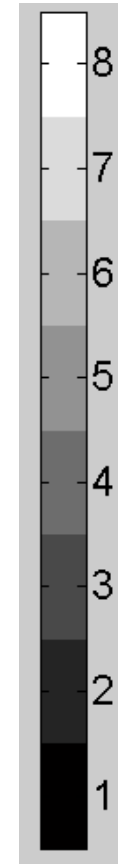
Choisir le nombre de niveaux de gris utilisés pour coder les valeurs d'intensité des pixels de l'image

Classiquement:

Image codée sur 8 bits: 256 valeurs possibles de 0 (noir) à 255 (blanc)



256 niveaux



8 niveaux

Quantification



256 niveaux



64 niveaux



32 niveaux



12 niveaux



8 niveaux



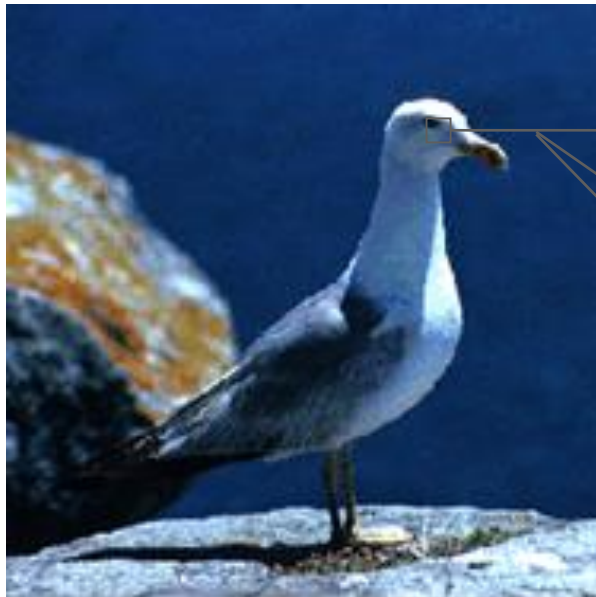
4 niveaux



2 niveaux

Une image couleur

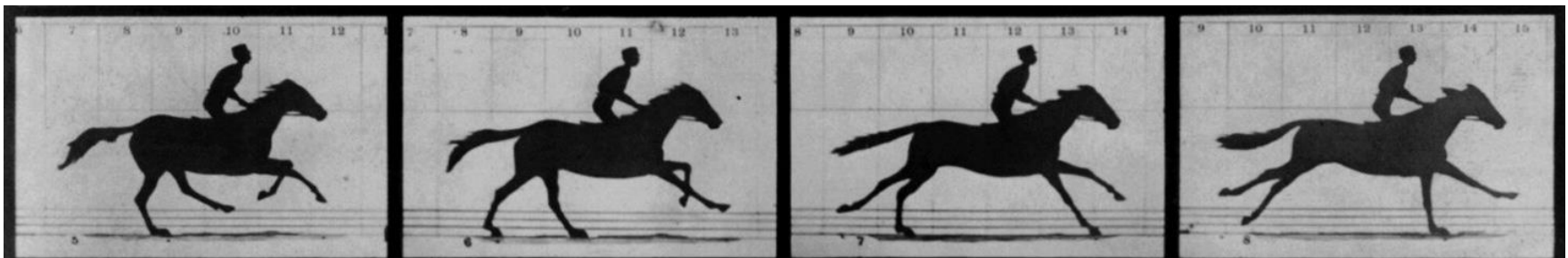
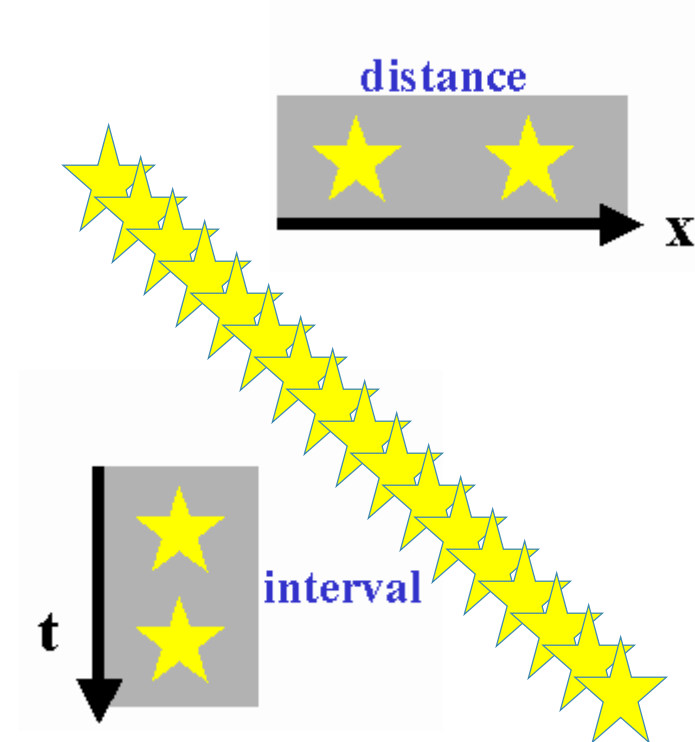
- Une image couleur comprend 3 plans (Rouge-Vert-Bleu). Les 3 plans ont la même résolution spatiale (même nombre de pixels).
- Un pixel possède donc 3 valeurs:
 - 1 valeur pour le Rouge
 - 1 valeur pour le Vert
 - 1 valeur pour Bleu



142	174	164	218	250	255	250	252	255	255
107	107	102	80	127	174	237	218	252	255
90	34	24	34	34	24	51	88	127	164
80	26	19	53	34	19	24	85	117	137
78	76	34	44	26	26	34	24	71	90
85	85	90	26	26	26	26	34	76	83
88	102	90	53	26	26	34	73	85	78
102	110	105	90	98	105	105	110	107	93
107	115	110	110	110	117	115	110	107	105
105	110	110	117	110	132	115	110	107	105
105	110	110	117	110	132	115	110	107	105
80	26	19	53	34	19	24	85	117	137
78	76	34	44	26	26	34	24	71	90
85	85	90	26	26	26	26	34	76	83
88	102	90	53	26	26	34	73	85	78
102	110	105	90	98	105	105	110	107	93
107	115	110	110	110	117	115	110	107	105
105	110	110	117	110	132	115	110	107	105
105	110	110	117	110	132	115	110	107	105
80	26	19	53	34	19	24	85	117	137
78	76	34	44	26	26	34	24	71	90
85	85	90	26	26	26	26	34	76	83
88	102	90	53	26	26	34	73	85	78
102	110	105	90	98	105	105	110	107	93
107	115	110	110	110	117	115	110	107	105
105	110	110	117	110	132	115	110	107	105
105	110	110	117	110	132	115	110	107	105
80	26	19	53	34	19	24	85	117	137
78	76	34	44	26	26	34	24	71	90
85	85	90	26	26	26	26	34	76	83
88	102	90	53	26	26	34	73	85	78
102	110	105	90	98	105	105	110	107	93
107	115	110	110	110	117	115	110	107	102
105	110	110	117	110	132	115	110	107	105
105	110	110	117	110	132	115	110	107	105

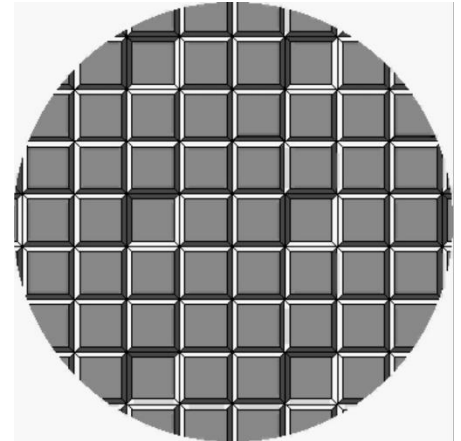
Une image en mouvement

- Mouvement = variation de position au cours du temps
 - $x = f(t)$
 - Une image en mouvement = variation de la quantification des pixels au cours du temps
-
- Un film est une succession d'images (24, 25 ou 30 images/frames par secondes)



Profondeur

- La perception de la profondeur se base sur:
 - Des indices monoculaires: ombres, occlusions, perspective, hauteur et taille relatives des objets et des gradients de texture.

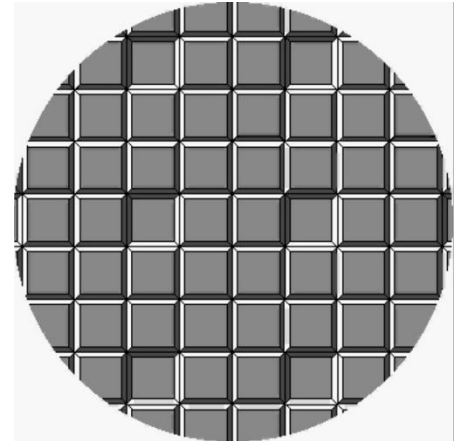


Profondeur



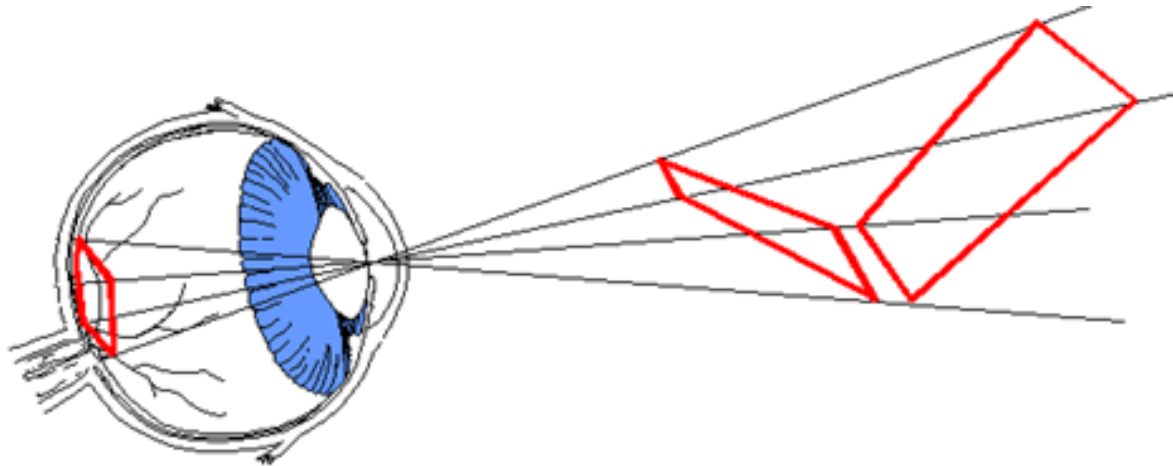
Profondeur

- La perception de la profondeur se base sur:
 - Des indices monoculaires: ombres, occlusions, perspective, hauteur et taille relatives des objets et des gradients de texture.
 - **Des indices binoculaires: décalage de l'information visuelle entre l'œil gauche et l'œil droit**
 - Des indices issus du système oculomoteur (vergence et accommodation)



Disparité binoculaire

- Une image rétinienne correspond à l'image de plusieurs objets différents dans l'environnement
- Si nous devons reconstruire une image 3D seulement à partir d'une image 2D rétinienne, il y aurait une infinité de solutions.

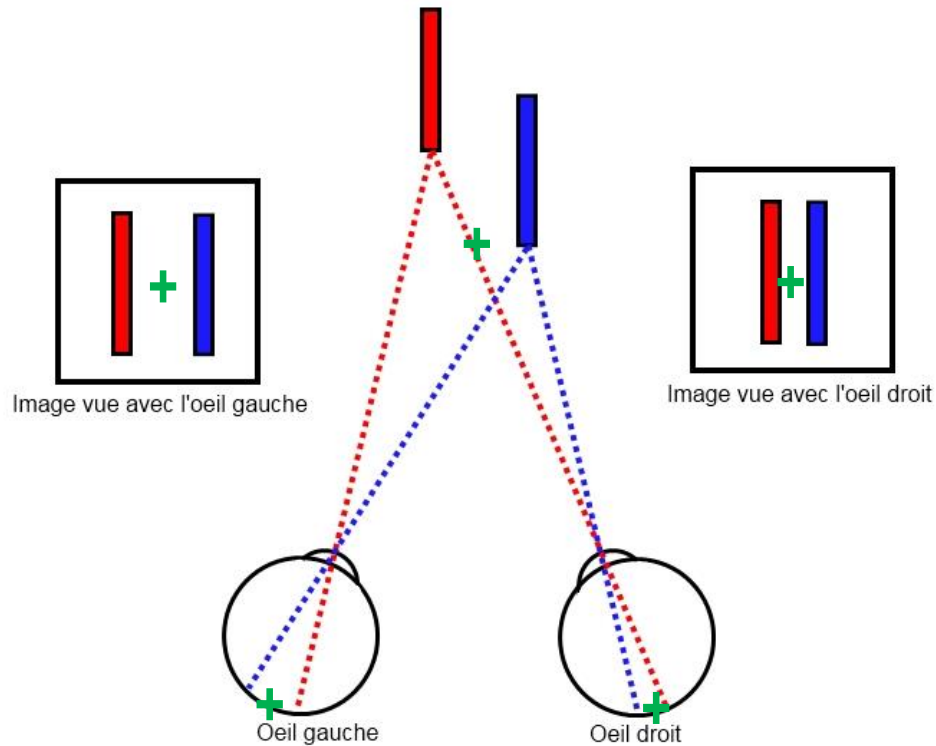


Solutions :

- Changer l'axe de vue : mouvements des objets ou de l'observateur
- **S'appuyer sur la disparité binoculaire (donne un indice de profondeur)**

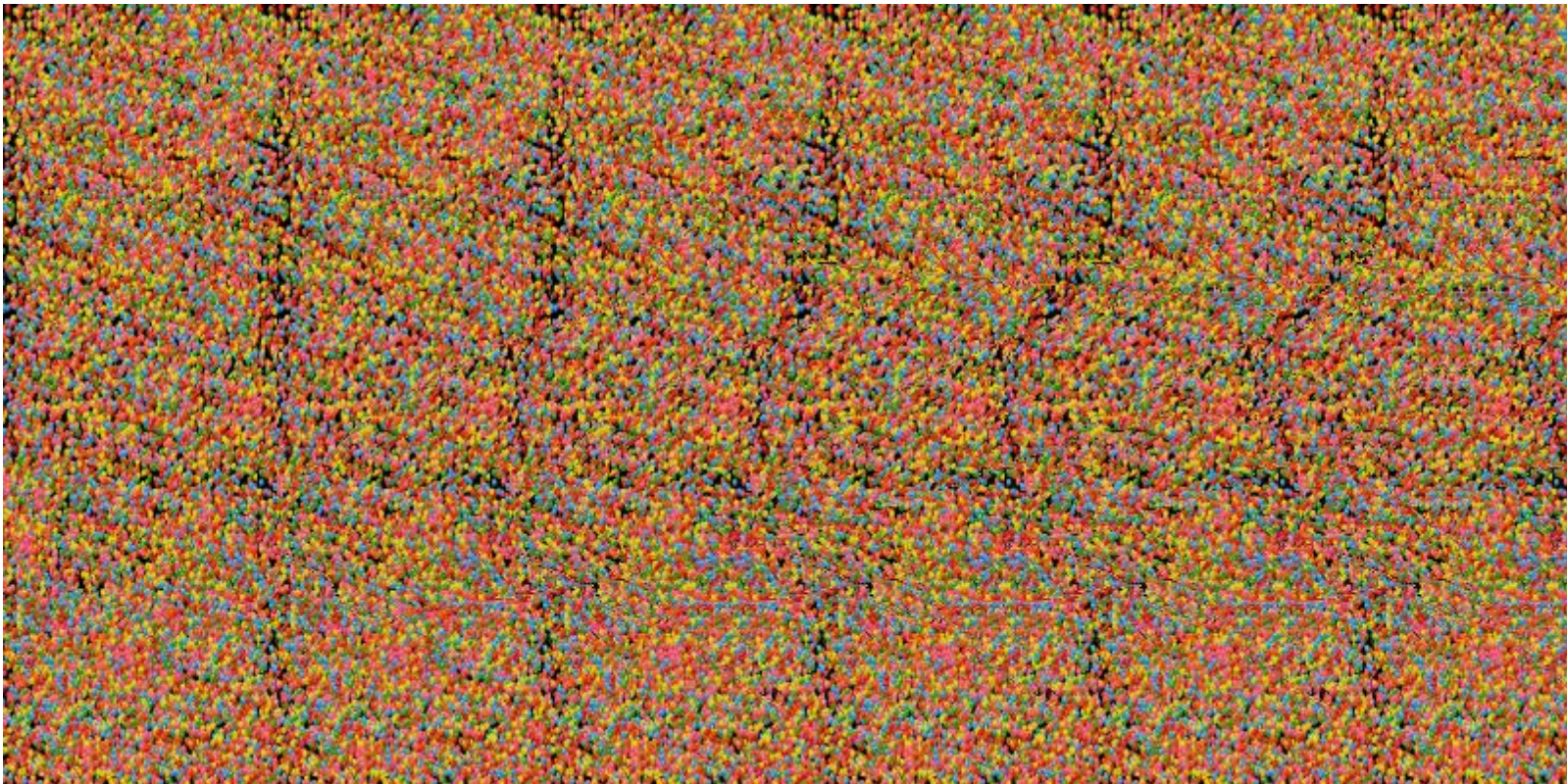
Disparité binoculaire

Il y a perception d'une profondeur entre deux objets lorsqu'il y a un décalage horizontal des images rétiniennes entre les deux yeux.



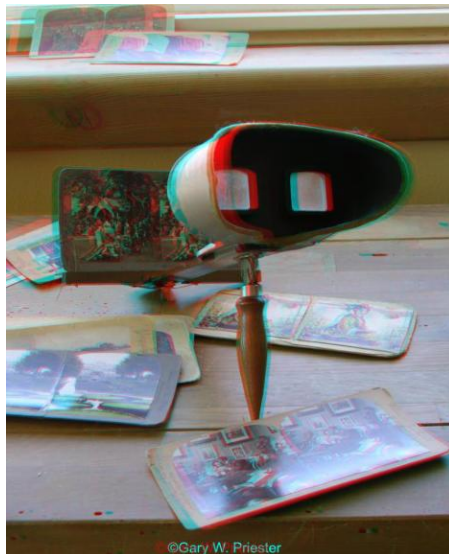
Illusion de la profondeur

- Ainsi, on peut simuler une perception de profondeur
 - En prenant un motif répétitif et en décalant légèrement un des motifs sur le plan horizontal = auto-stéréogramme
 - recrée l'ambiguïté existant entre un décalage horizontal et un décalage du à la profondeur



Illusion de la profondeur

- Ainsi, on peut simuler une perception de profondeur
 - En prenant un motif répétitif et en décalant légèrement un des motifs sur le plan horizontal = auto-stéréogramme
 - recrée l'ambiguïté existant entre un décalage horizontal et un décalage du à la profondeur
 - En prenant deux photos avec un point de vue légèrement décalé et en combinant ces photos en utilisant les différents plans d'une image (e.g., couleur) → utiliser un filtre (e.g., chromatique) = anaglyphes



Pour créer un anaglyphe...

Pour faire un anaglyphe, il suffit de construire une image dont les plans **ROUGE** et **CYAN** (**BLEU+VERT**) sont constitués de deux points de vue différents d'une même image → permet de projeter à chaque oeil une version décalée de l'image qui sera interprétée comme reflétant une profondeur.

```
% Removes green and blue from the left eye image
```

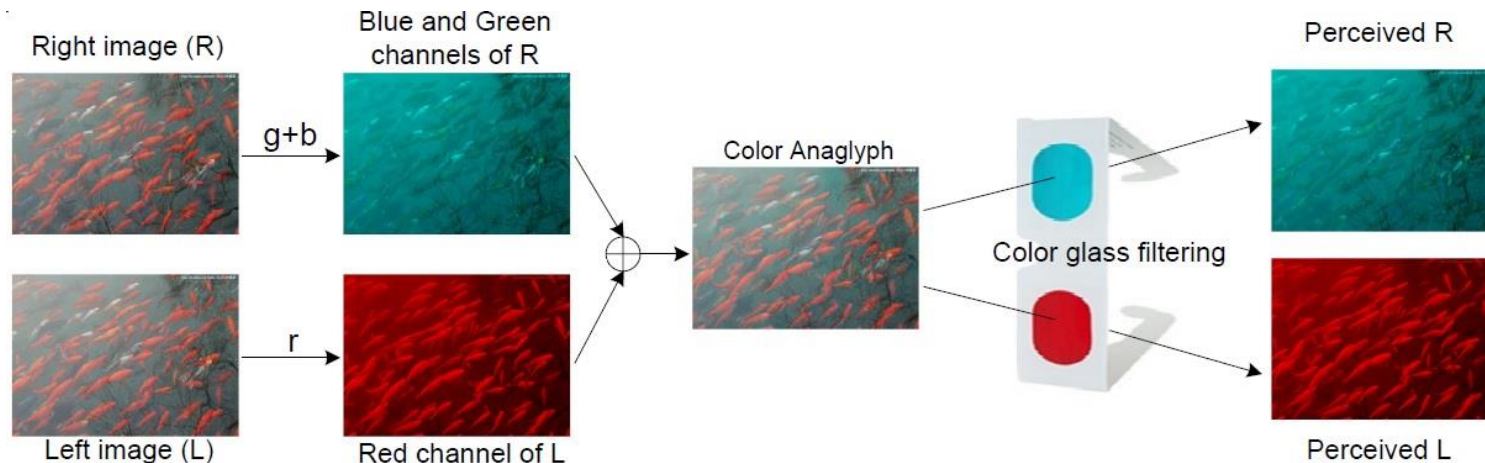
```
leftEyeImage(:, :, 2:3) = 0;
```

```
% Removes red from the right eye image
```

```
rightEyeImage(:, :, 1) = 0;
```

```
% Combines the two to produce the finished anaglyph
```

```
anaglyph = leftEyeImage + rightEyeImage;
```

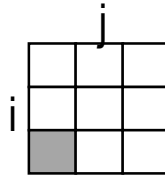


Synthèse: Un pixel

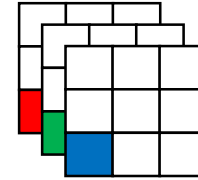
- Pixel = « picture element »
- Défini par
 - Sa position $x = (i,j)$
 - Sa taille (résolution) et son codage (Intensité ou Amplitude)

- Dimensions de représentation:

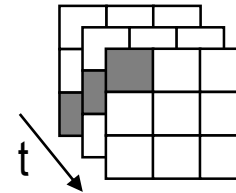
- **Image NB** (luminance) en 2D



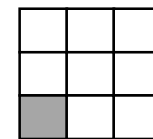
- **Couleur**: une dimension supplémentaire = 3 canaux de couleur



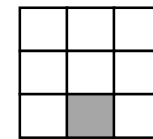
- **Mouvement**: une dimension supplémentaire = temps



- **Profondeur**: une dimension supplémentaire = OD/OG



OG



OD

Synthèse: Une scène

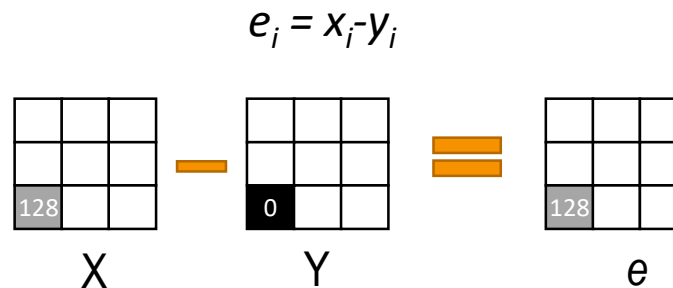
- Un ensemble de pixels dans un espace à plusieurs dimensions (spatiales, couleur, temps, etc)
- Comment définir des catégories de scènes?
 - Quelles propriétés sont communes aux exemplaires d'une catégorie?
 - Quelles propriétés sont différentes entre différentes catégories?
- Notion de similitude et de différence → calcul de distance



Comment déterminer les ressemblances/différences entre deux scènes?

Exemple: le MSE

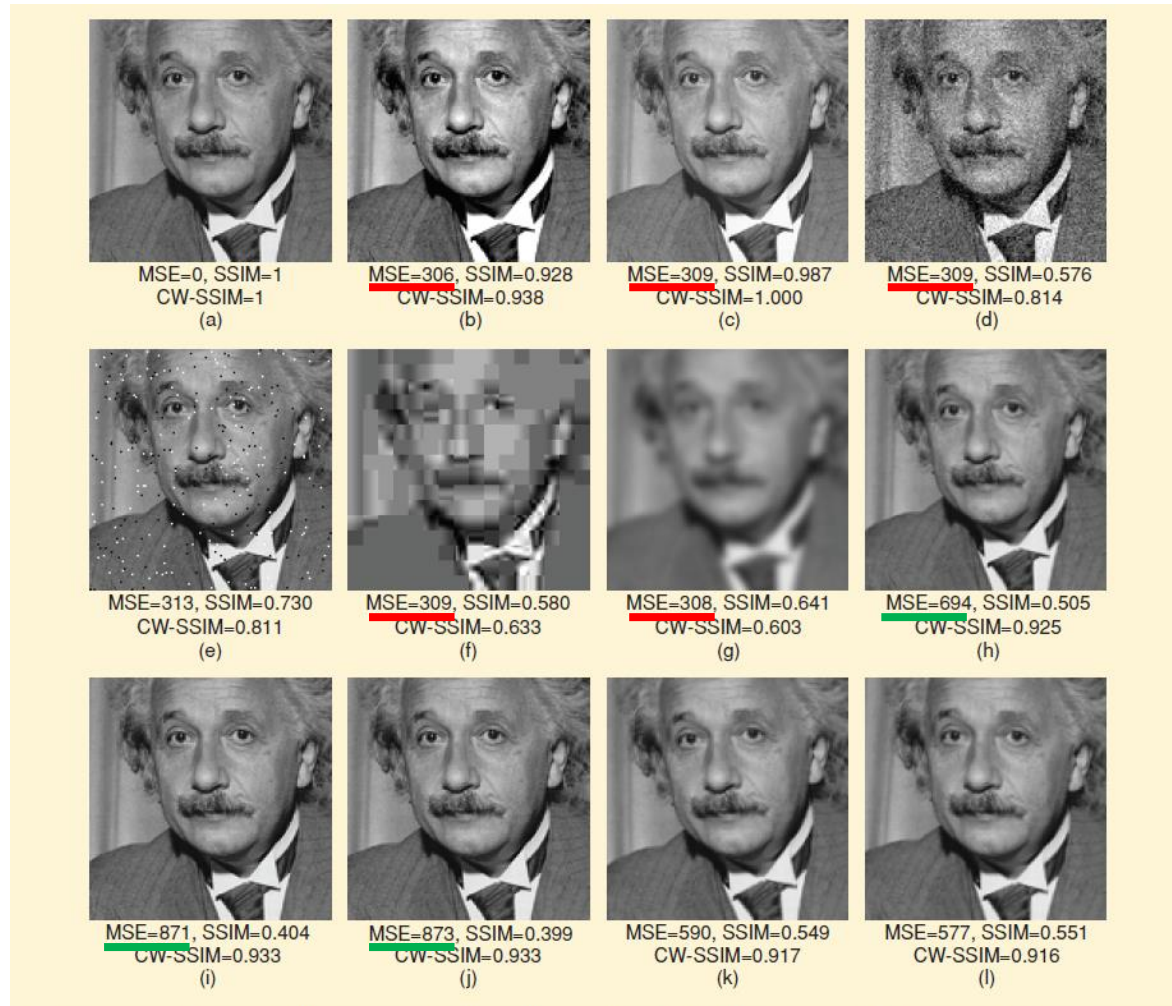
- La mesure de distance (e) la plus simple est la différence entre deux images (X et Y):



- Le calcul du **MSE** (Mean Squared Error) résume cela en une valeur:

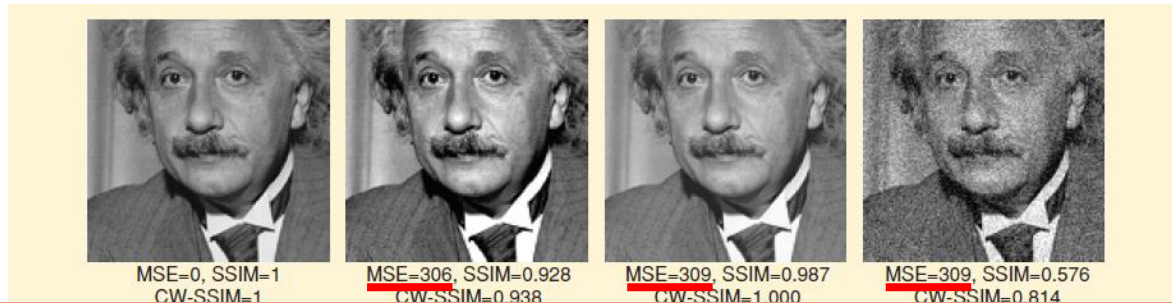
$$\text{MSE}(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2.$$

Comment déterminer les ressemblances/différences entre deux scènes?



[FIG2] Comparison of image fidelity measures for "Einstein" image altered with different types of distortions. (a) Reference image. (b) Mean contrast stretch. (c) Luminance shift. (d) Gaussian noise contamination. (e) Impulsive noise contamination. (f) JPEG compression. (g) Blurring. (h) Spatial scaling (zooming out). (i) Spatial shift (to the right). (j) Spatial shift (to the left). (k) Rotation (counter-clockwise). (l) Rotation (clockwise).

Comment déterminer les ressemblances/différences entre deux scènes?



Problème: Ce type de mesure rend peu compte de la similarité perceptive humaine

→ Nécessité de proposer des métriques plus proches de la perception visuelle

*Le pixel est-il la bonne unité de description?
L'espace de l'image est-il le plus approprié?*

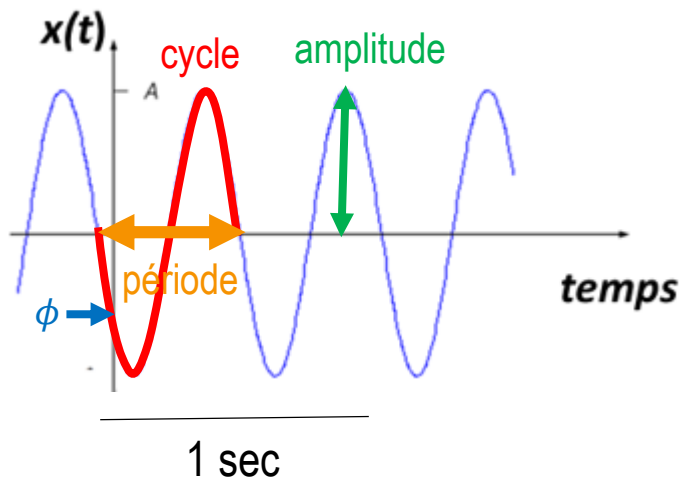


Un autre espace de représentation: l'espace de Fourier

(counter-clockwise). (I) Rotation (clockwise).

Rappels sur les signaux périodiques

- Onde sinusoïdale en fonction du temps:



$$T = 0.5 \text{ secondes}$$

f = fréquence (nombre de cycles / unité de temps)

$$f = 2/\text{sec} = 2\text{Hz}$$

$$T = 1/f = 1/2 = 0.5$$

$$x(t) = A \sin(2\pi f t + \phi)$$

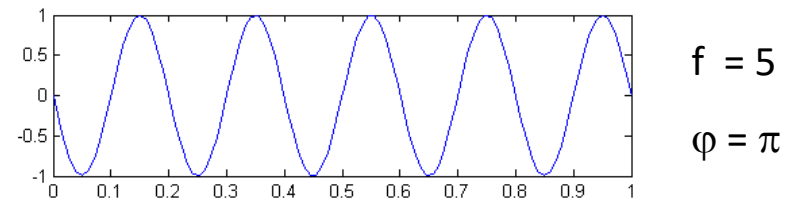
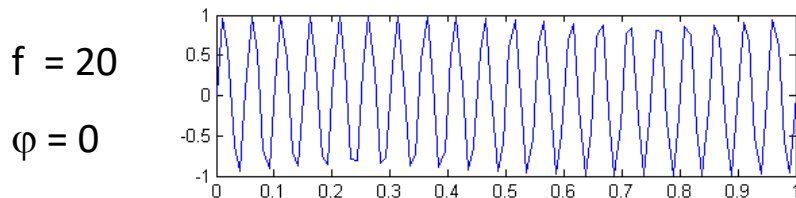
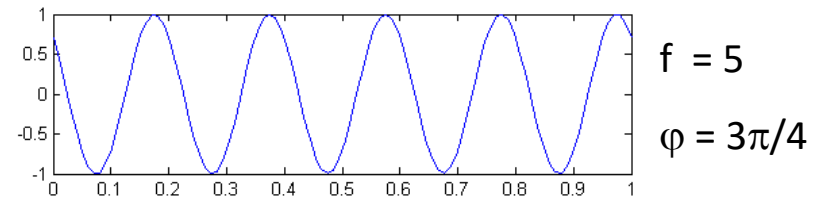
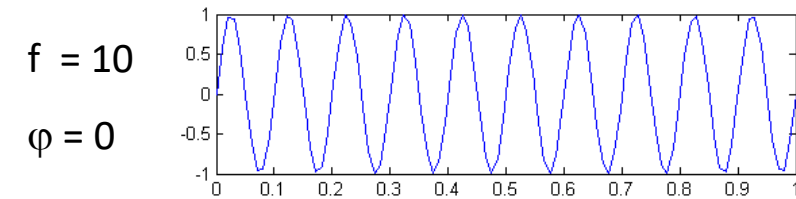
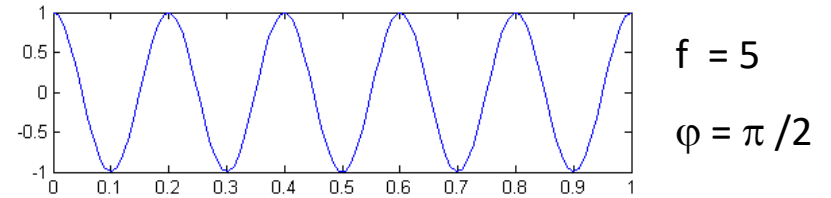
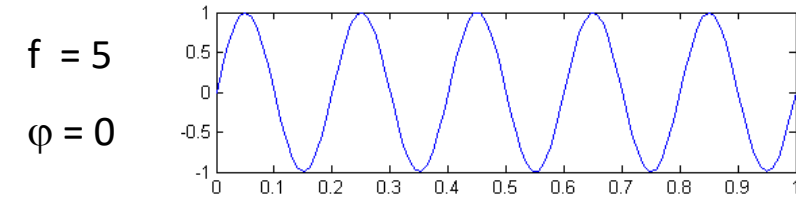
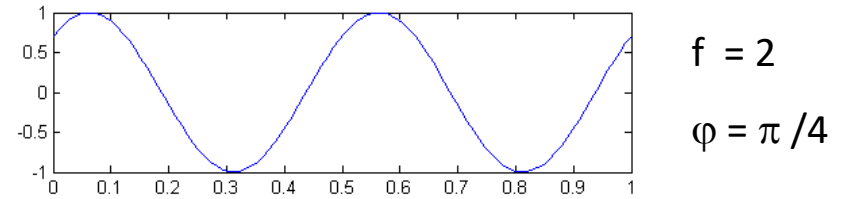
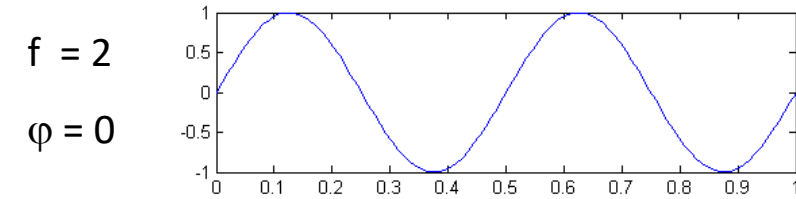
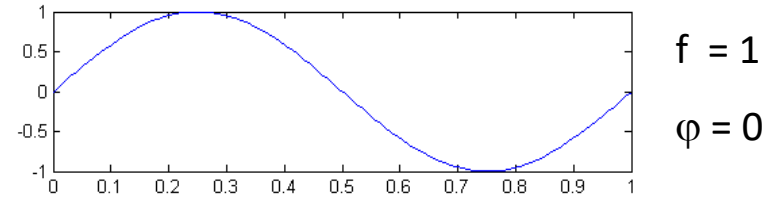
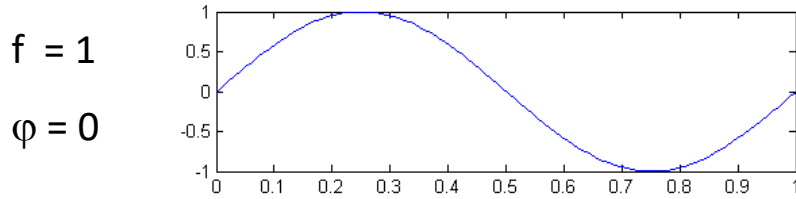
- $x(t)$ est un signal dont l'amplitude (A) varie de façon sinusoïdale en fonction du temps selon:

$$x(t) = A \sin\left(\frac{2\pi}{T} t + \phi\right)$$

Avec:

- A = amplitude du signal (distance maximum par rapport à 0)
- T = période (en secondes) = temps pour effectuer un cycle
- ϕ = déphasage (en radians) = où l'onde commence par rapport au cycle d'oscillation

Rappels sur les signaux périodiques



➔ Signaux différents mais variations similaires

La transformée de Fourier

Soit un signal certain $x(t)$, sa transformée de Fourier (TF) est une fonction complexe de la variable réelle f définie par:

$$\underbrace{X(f)}_{\text{Signal fréquentiel}} = \int_{-\infty}^{+\infty} \underbrace{x(t)}_{\text{Signal temporel}} \exp(-j2\pi ft) dt$$

Avec, le module:

$$|X(f)| = \sqrt{X(f)X^*(f)} = \sqrt{\text{Re}(X(f))^2 + \text{Im}(X(f))^2}$$

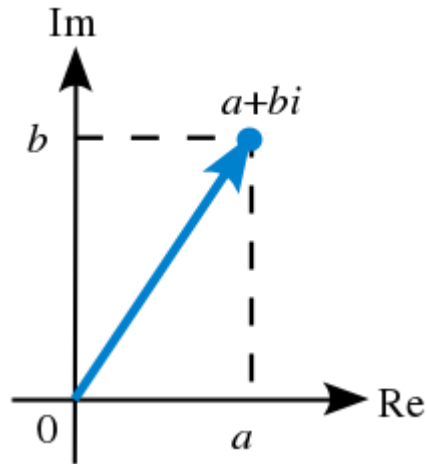
et la phase:

$$\varphi(X(f)) = \text{arctg} \left(\frac{\text{Im}(X(f))}{\text{Re}(X(f))} \right)$$

On parle également de spectre d'amplitude et de spectre de phase

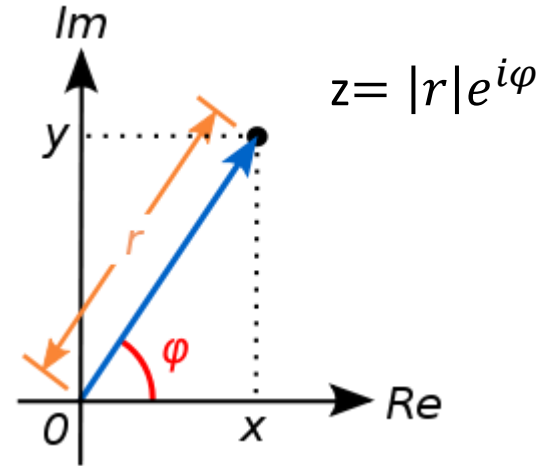
Rq: la TF est une extension, pour les fonctions non périodiques, du développement en série de Fourier des fonctions périodiques.

Nombre complexe



Forme cartésienne

Un nombre complexe peut être représenté visuellement par une paire de nombres (a, b) formant un **vecteur** dans le plan complexe, avec Re la partie réelle et Im la partie imaginaire où $i^2 = -1$

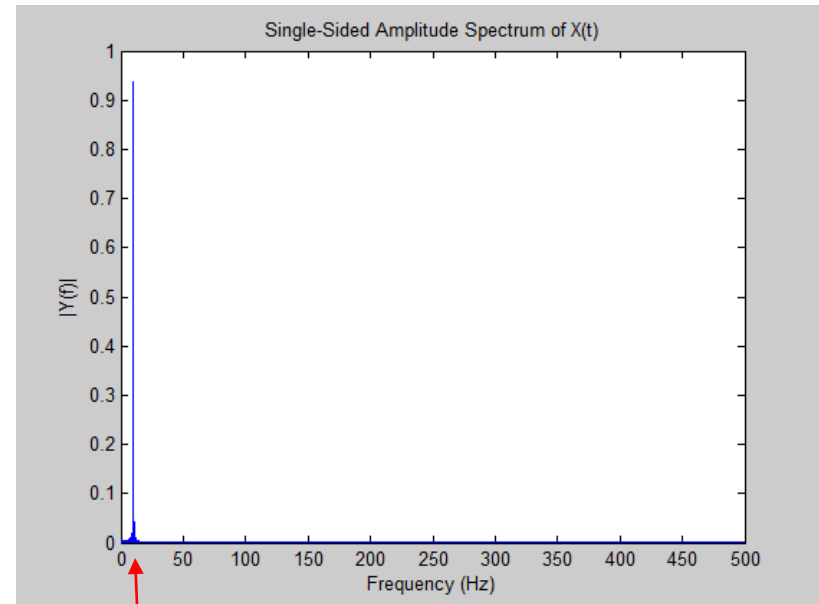
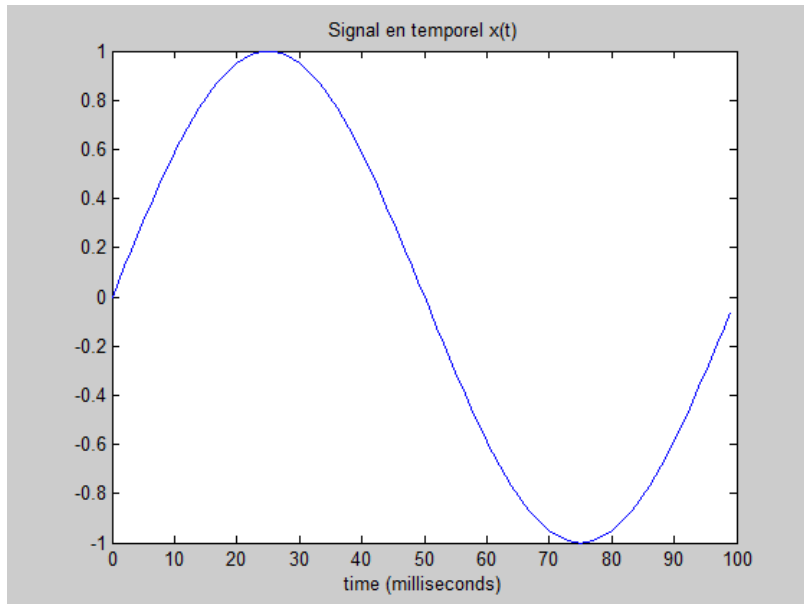


Forme géométrique

Une autre façon de représenter ce nombre est sa forme « polaire »: avec r (**module ou magnitude**) = valeur absolue du nombre complexe (distance par rapport à l'origine) et ϕ (**phase**) = angle avec l'axe Re

La transformée de Fourier

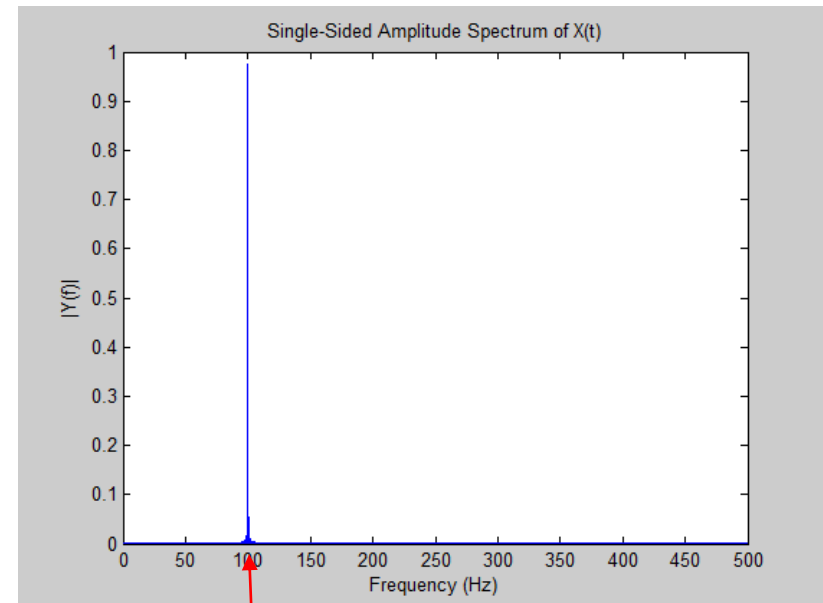
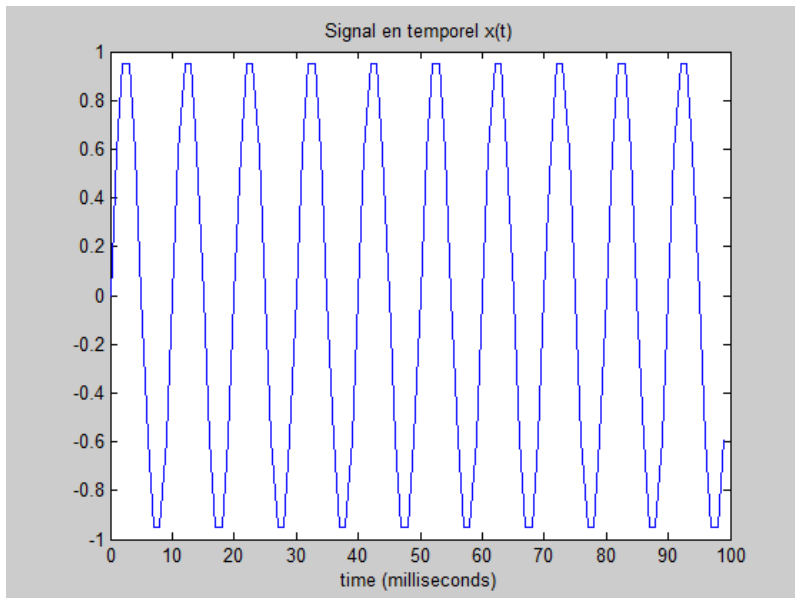
La transformée de Fourier permet une analyse harmonique du signal via une représentation spectrale de celui-ci: elle exprime la **répartition de l'amplitude et de la phase** de l'énergie d'un signal **en fonction des fréquences**.



$$F = 1 \text{ cycle} / 100 \text{ ms} = 10 \text{ cycles} / \text{sec} (10\text{Hz})$$

La transformée de Fourier

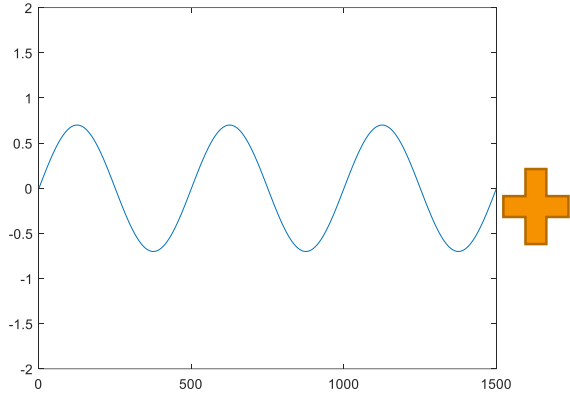
La transformée de Fourier permet une analyse harmonique du signal via une représentation spectrale de celui-ci: elle exprime la **répartition de l'amplitude et de la phase de l'énergie d'un signal en fonction des fréquences.**



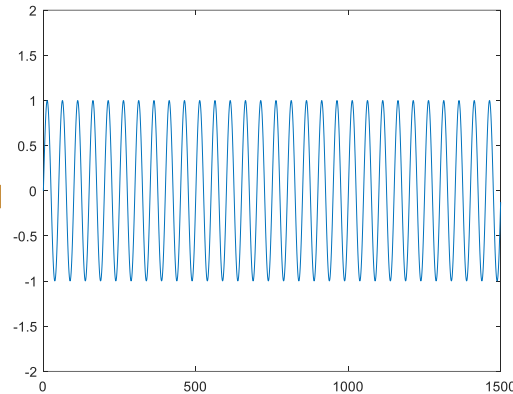
$$F = 10 \text{ cycles} / 100 \text{ ms} = 100 \text{ cycles} / \text{sec} (100\text{Hz})$$

La transformée de Fourier

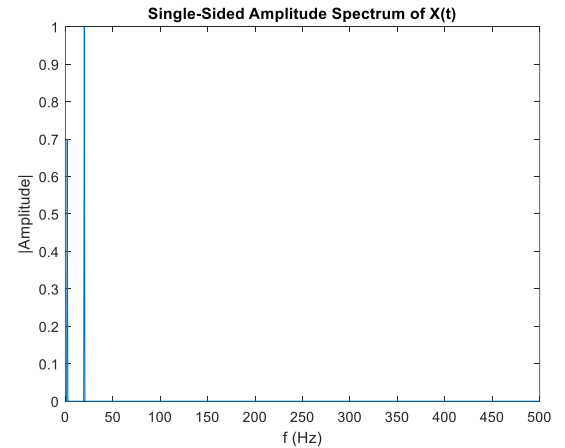
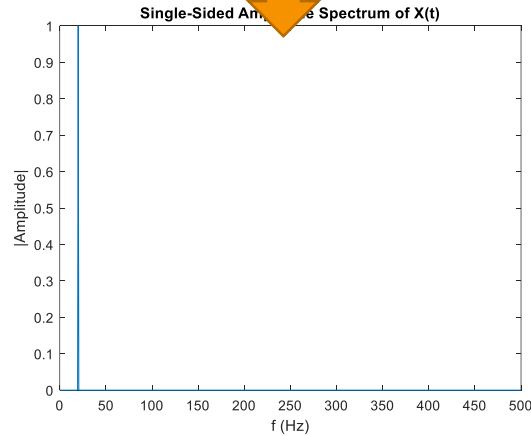
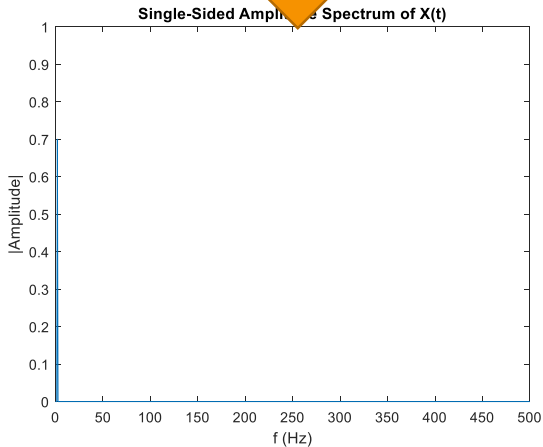
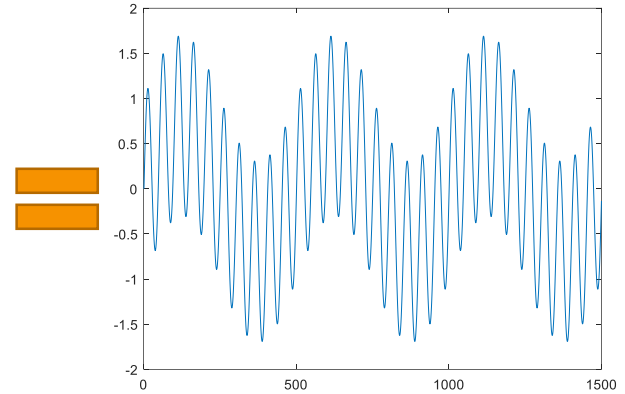
Un signal est décomposable en une somme de sinusoïdes à différentes fréquences:



S1: $f = 2\text{Hz}$, $A = 0,7$

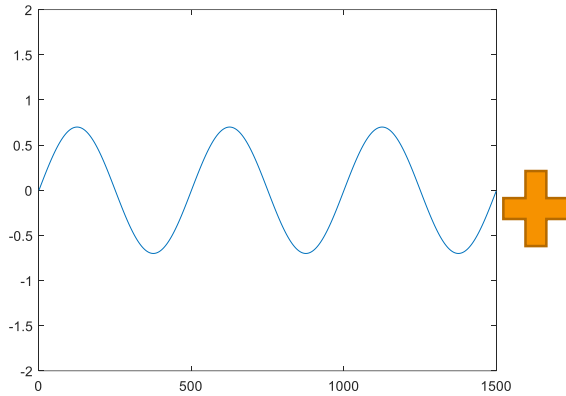


S2: $f = 20\text{Hz}$, $A = 1$

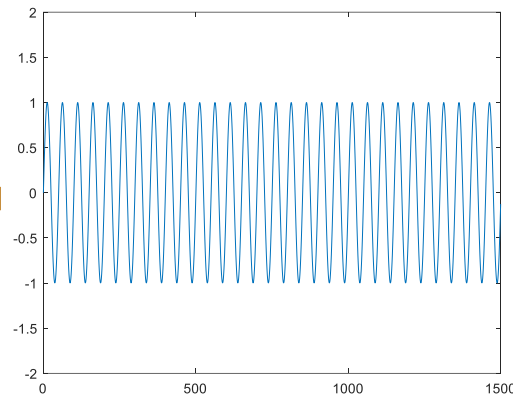


La transformée de Fourier

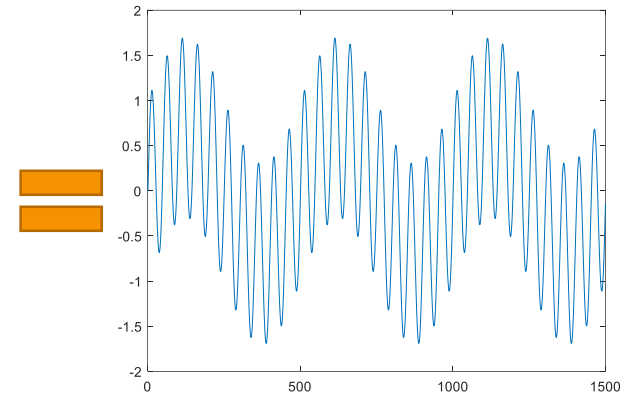
Un signal est décomposable en une somme de sinusoides à différentes fréquences:



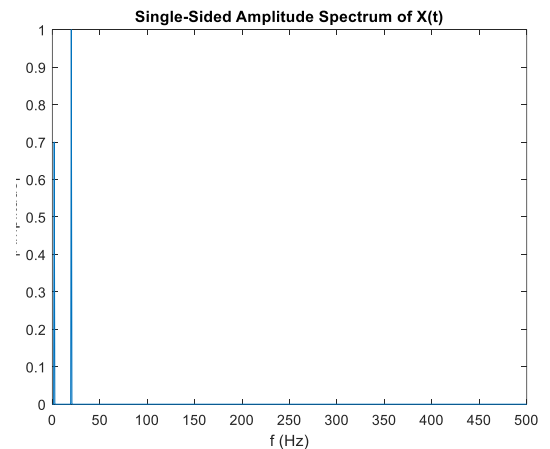
S1: $f = 2\text{Hz}$, $A = 0,7$



S2: $f = 20\text{Hz}$, $A = 1$



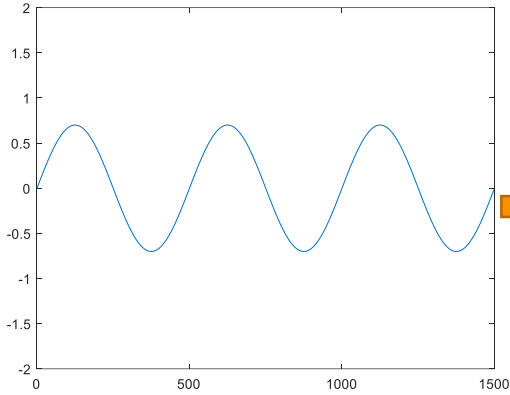
TF
↓



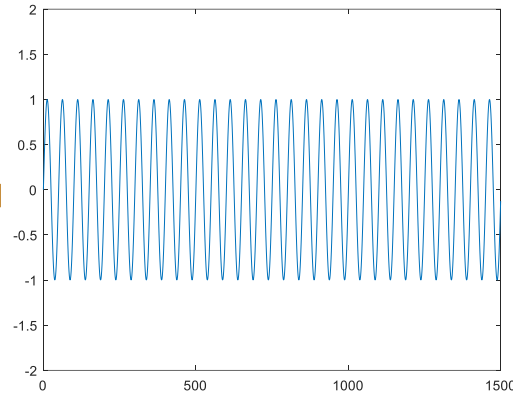
→ A partir du spectre on peut facilement analyser les fréquences qui composent un signal et éventuellement supprimer sélectivement certaines composantes du signal...

La transformée de Fourier

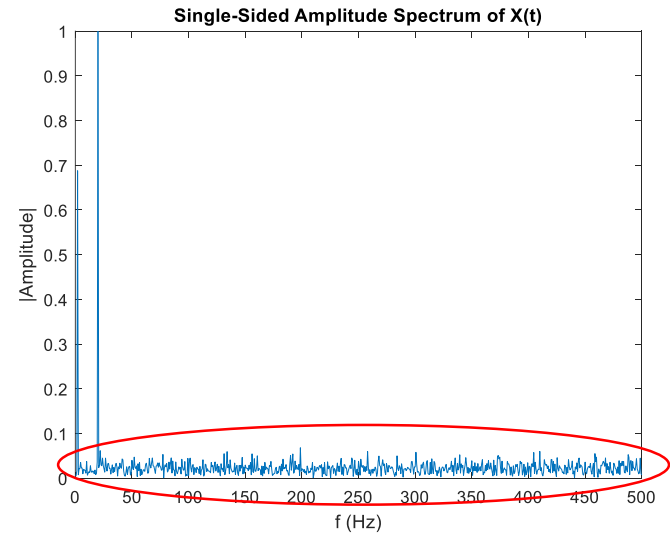
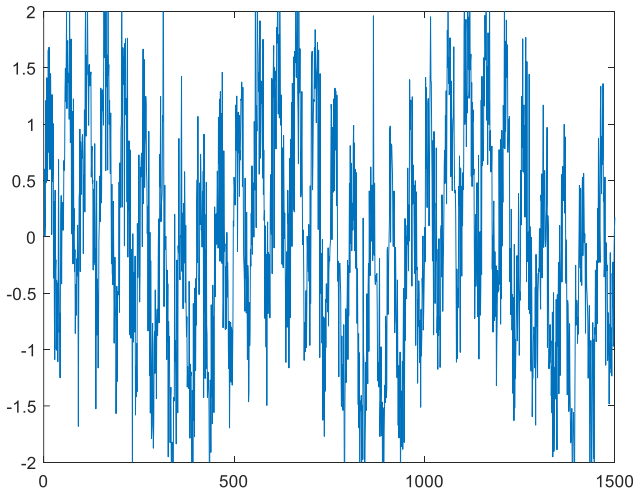
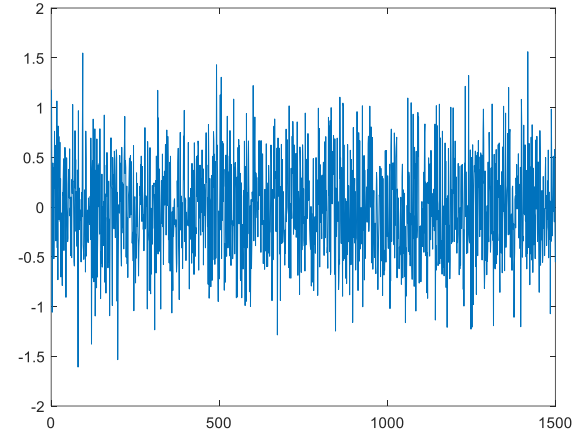
Exemple d'un signal bruité:



S1: $f = 2\text{Hz}$, $A = 0,7$



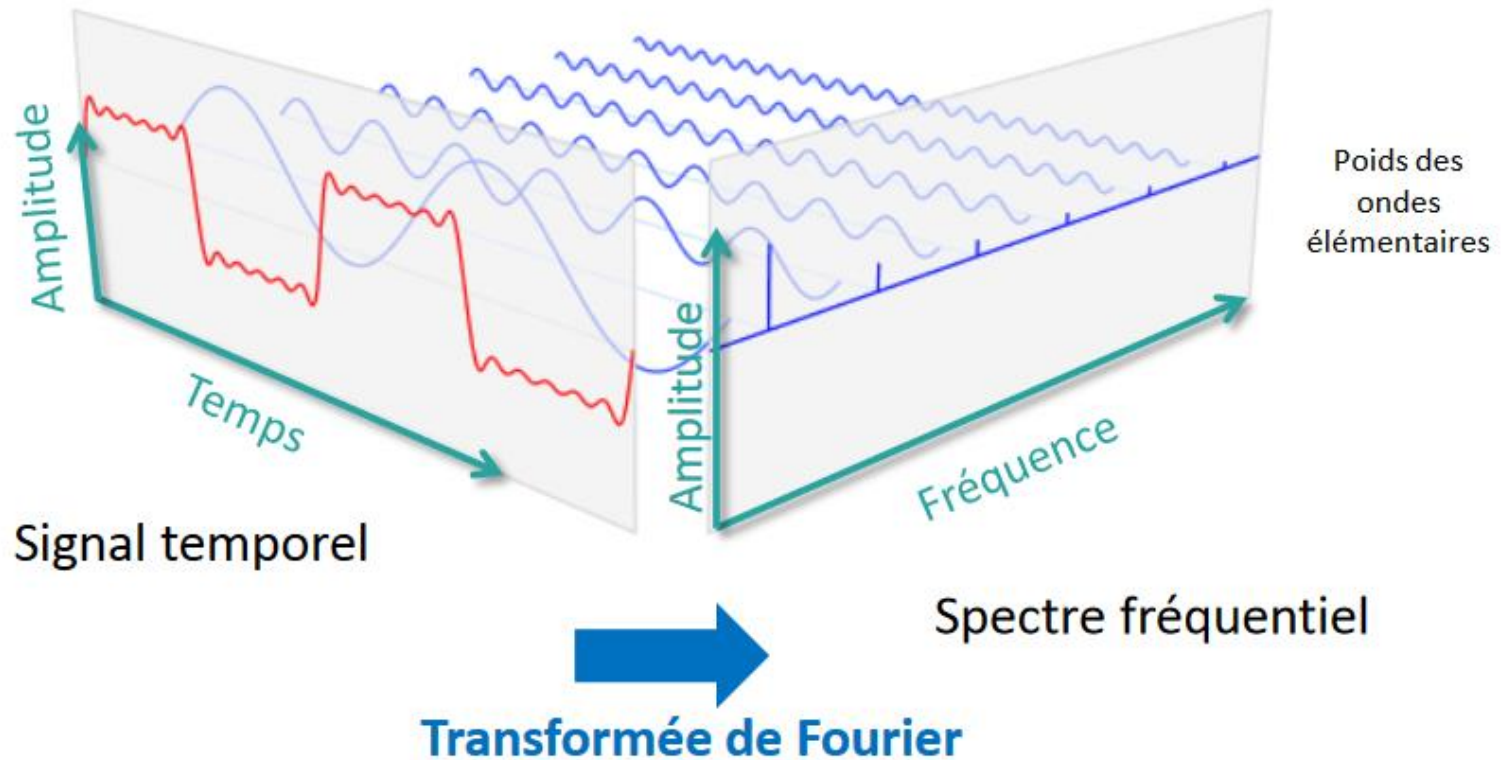
S2: $f = 20\text{Hz}$, $A = 1$



Résumé

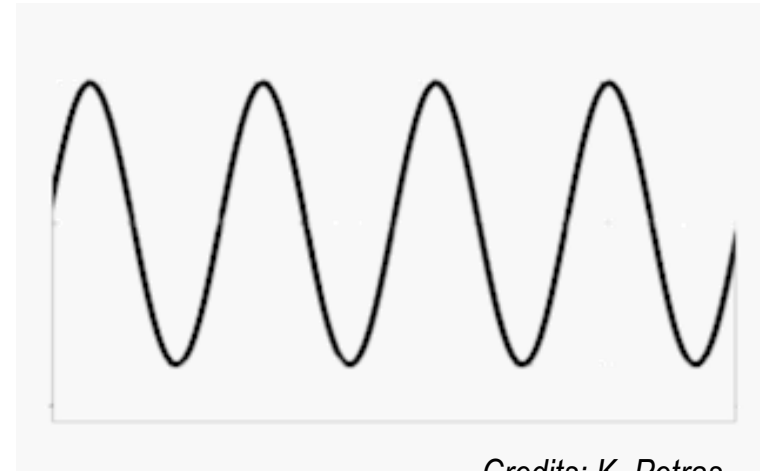
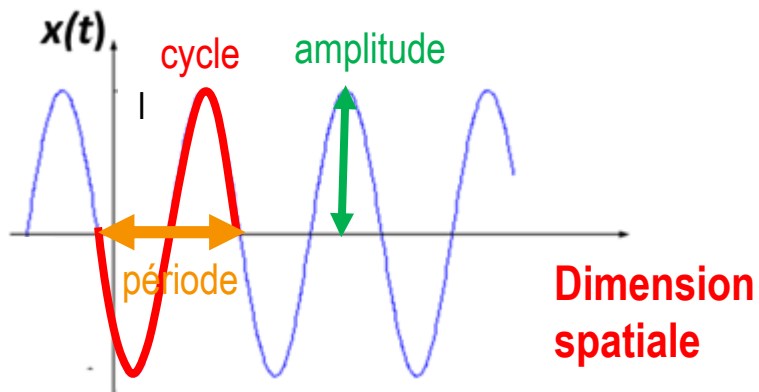


Résumé



Application à une image

- Exemple: Forme sinusoïdale des variations de luminance

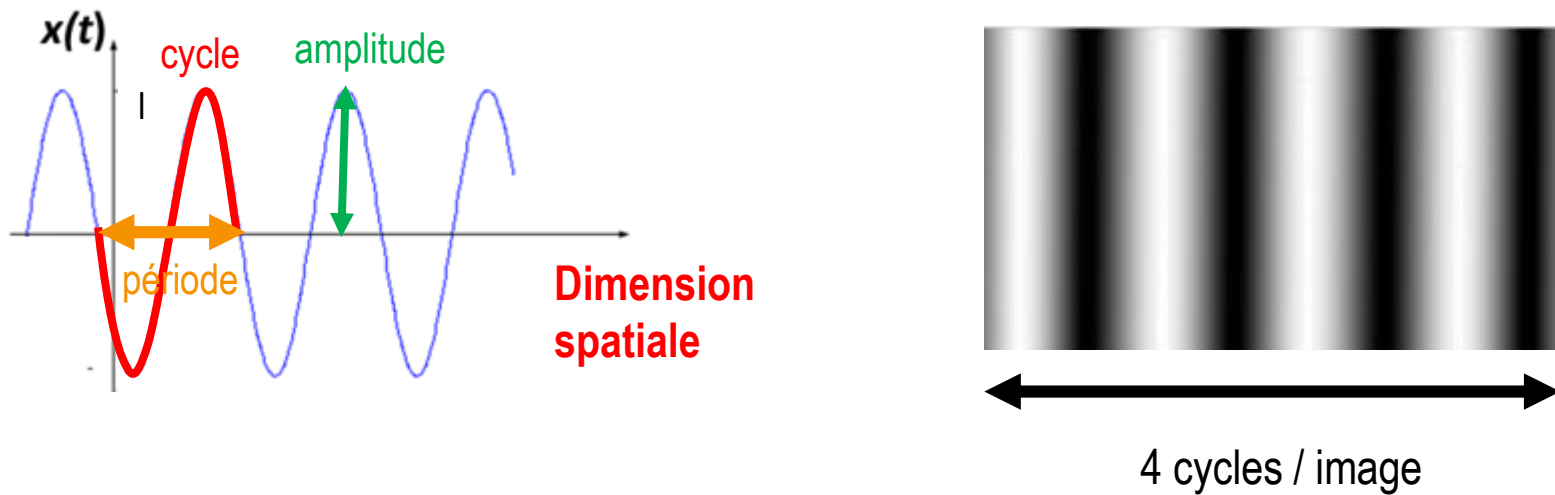


Ici:

- Amplitude exprimée par l'intensité des pixels (luminance)

Application à une image

- Exemple: Forme sinusoïdale des variations de luminance

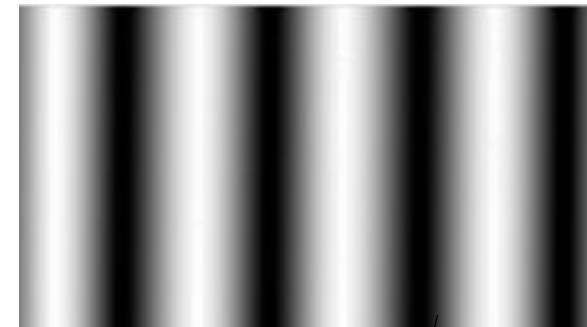
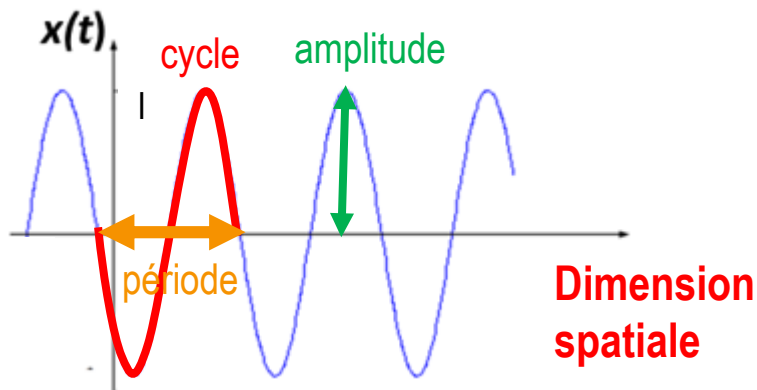


Ici:

- Amplitude exprimée par l'intensité des pixels (luminance)
- Fréquence spatiale exprimée en cycles / image (référentiel de l'image)

Application à une image

- Exemple: Forme sinusoïdale des variations de luminance



2 cycles / degrés

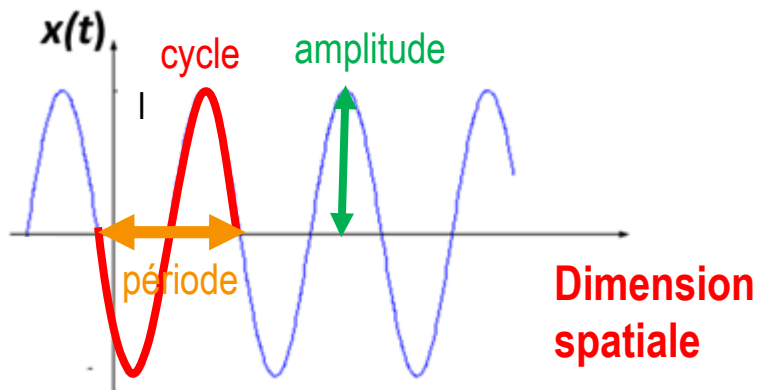
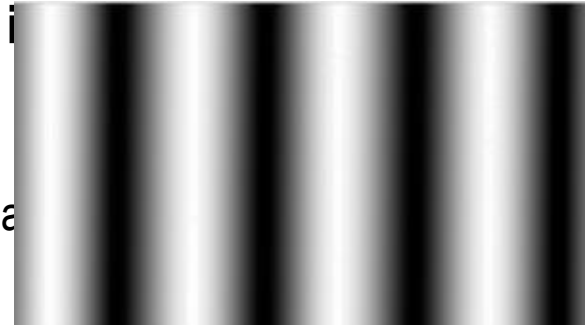


Ici:

- Amplitude exprimée par l'intensité des pixels (luminance)
- Fréquence spatiale exprimée en cycles / image (référentiel de l'image)
- ...Ou en cycles / degré d'angle visuel (référentiel de l'observateur)

Application à une image

- Exemple: Forme sinusoïdale des variations de luminosité



3 cycles / degrés

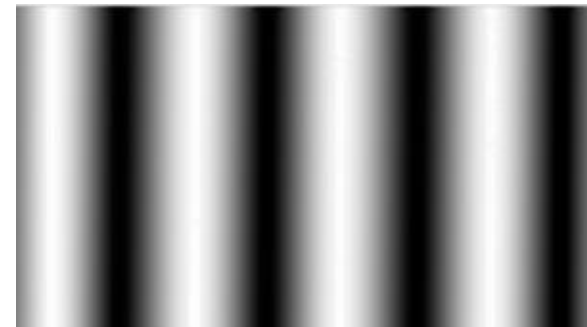
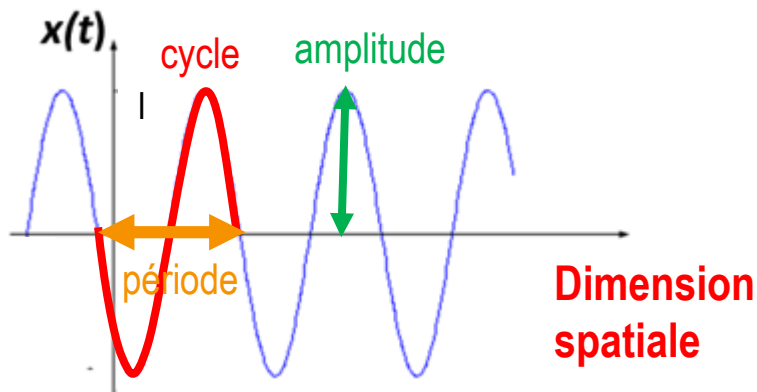
Ici:

- Amplitude exprimée par l'intensité des pixels (luminance)
- Fréquence spatiale exprimée en cycles / image (référentiel de l'image)
- ...Ou en cycles / degré d'angle visuel (référentiel de l'observateur)

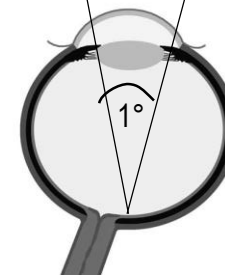


Application à une image

- Exemple: Forme sinusoïdale des variations de luminance



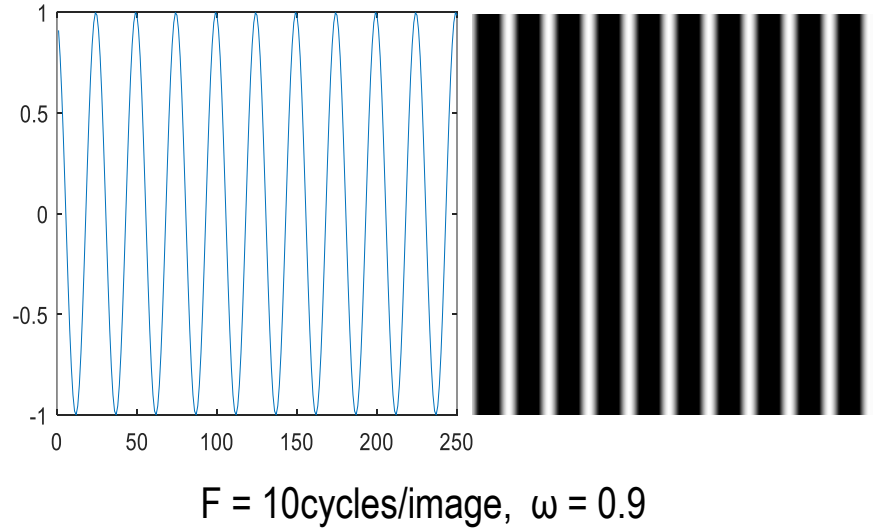
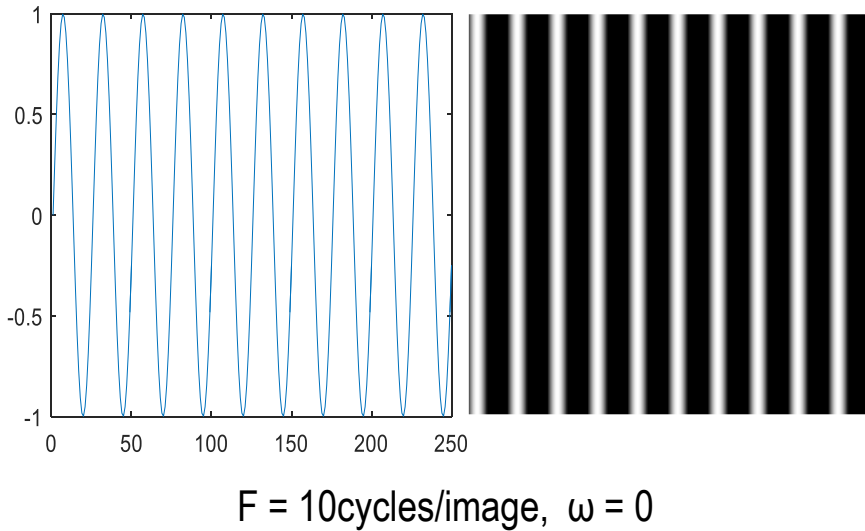
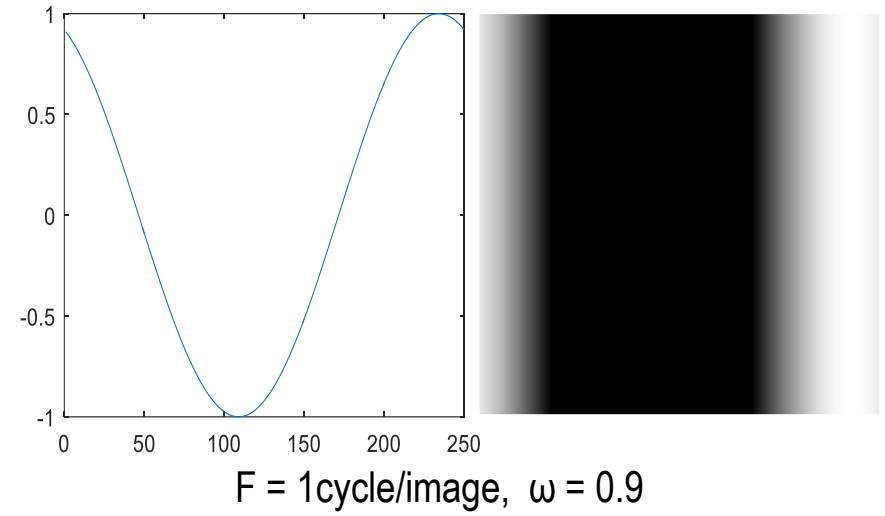
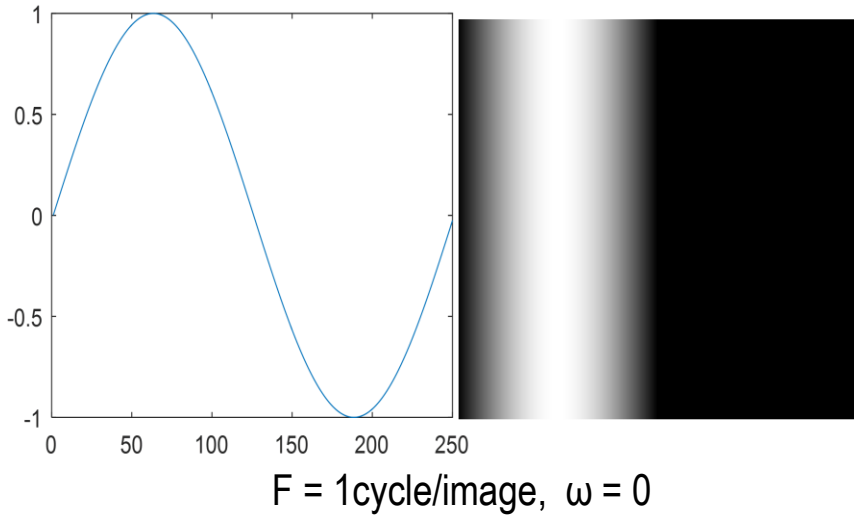
1 cycle / degrés



Ici:

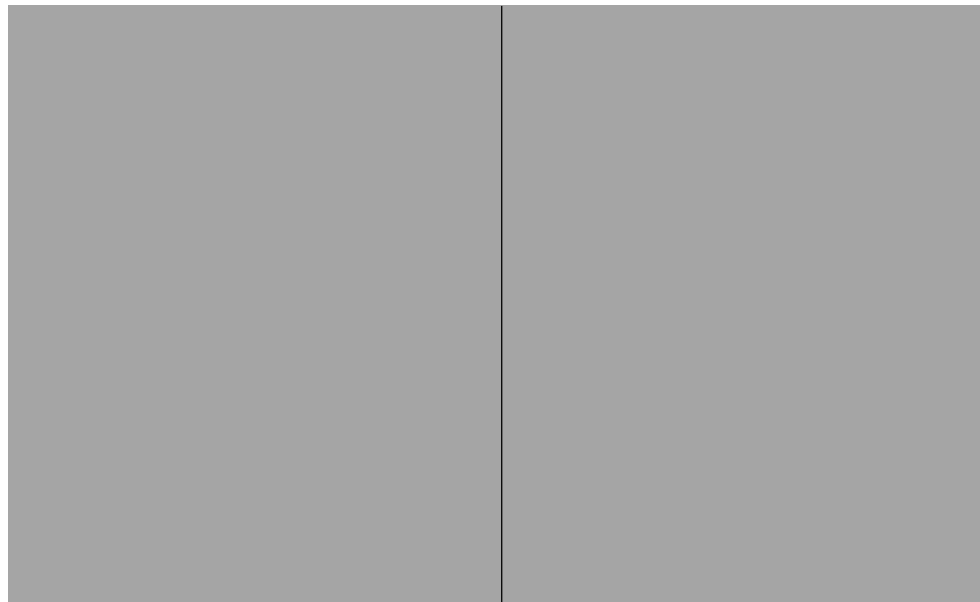
- Amplitude exprimée par l'intensité des pixels (luminance)
- Fréquence spatiale exprimée en cycles / image (référentiel de l'image)
- ...Ou en cycles / degré d'angle visuel (référentiel de l'observateur)

Application à une image



Application à une image

- De la même façon qu'un signal temporel complexe peut être décomposé en une somme de sinusoïdes, les variations spatiales de luminance dans une image complexe peuvent également être décomposées en une somme de sinusoïdes de différentes **fréquences spatiales** à **différentes orientations**

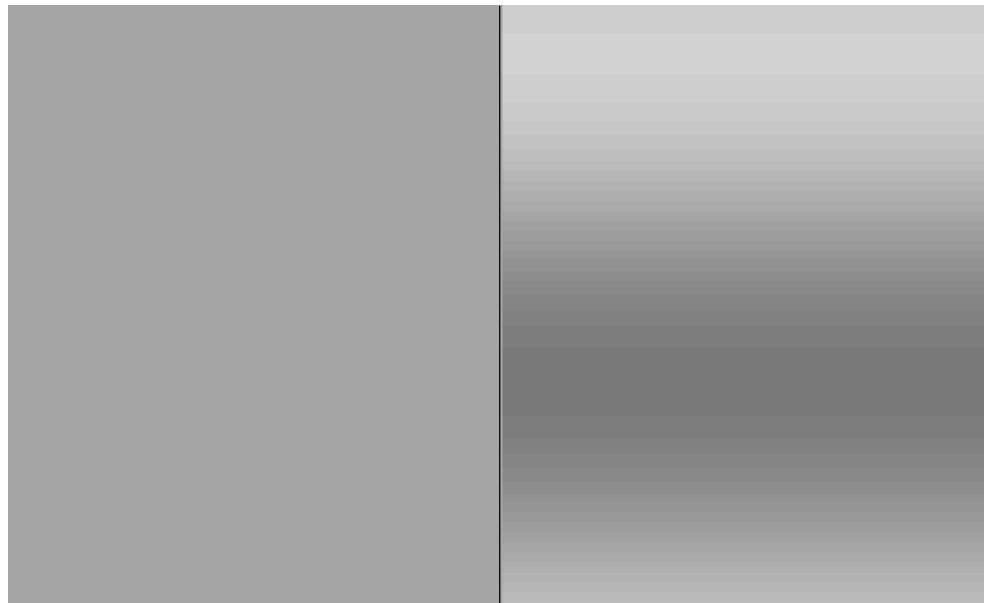


Sinusoïde

Somme de Sinusoïdes

Application à une image

- De la même façon qu'un signal temporel complexe peut être décomposé en une somme de sinusoïdes, les variations spatiales de luminance dans une image complexe peuvent également être décomposées en une somme de sinusoïdes de différentes **fréquences spatiales** à **différentes orientations**



Sinusoïde

Somme de Sinusoïdes

Application à une image

- De la même façon qu'un signal temporel complexe peut être décomposé en une somme de sinusoïdes, les variations spatiales de luminance dans une image complexe peuvent également être décomposées en une somme de sinusoïdes de différentes **fréquences spatiales** à **différentes orientations**

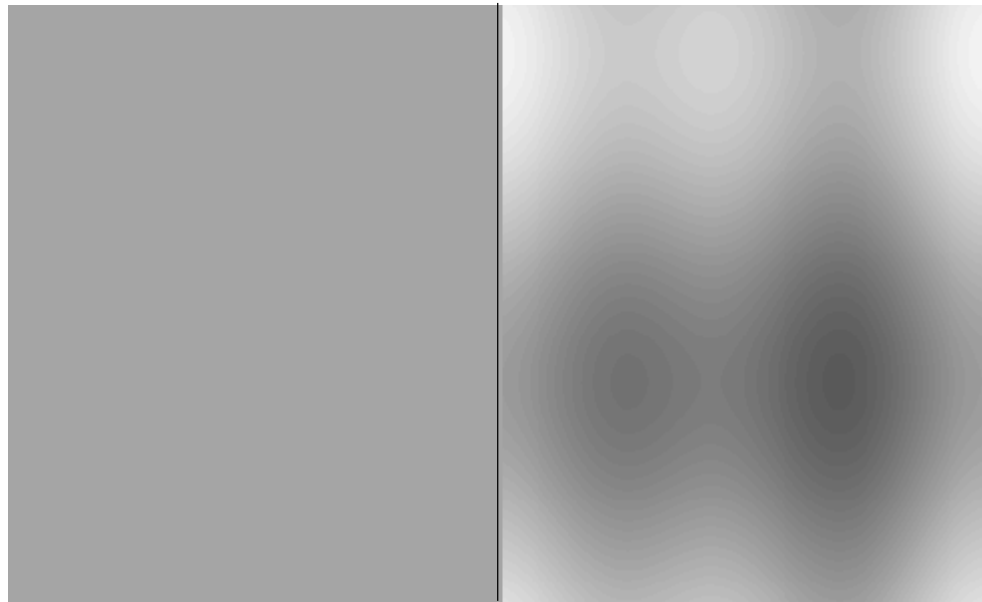


Sinusoïde

Somme de Sinusoïdes

Application à une image

- De la même façon qu'un signal temporel complexe peut être décomposé en une somme de sinusoïdes, les variations spatiales de luminance dans une image complexe peuvent également être décomposées en une somme de sinusoïdes de différentes **fréquences spatiales** à **différentes orientations**

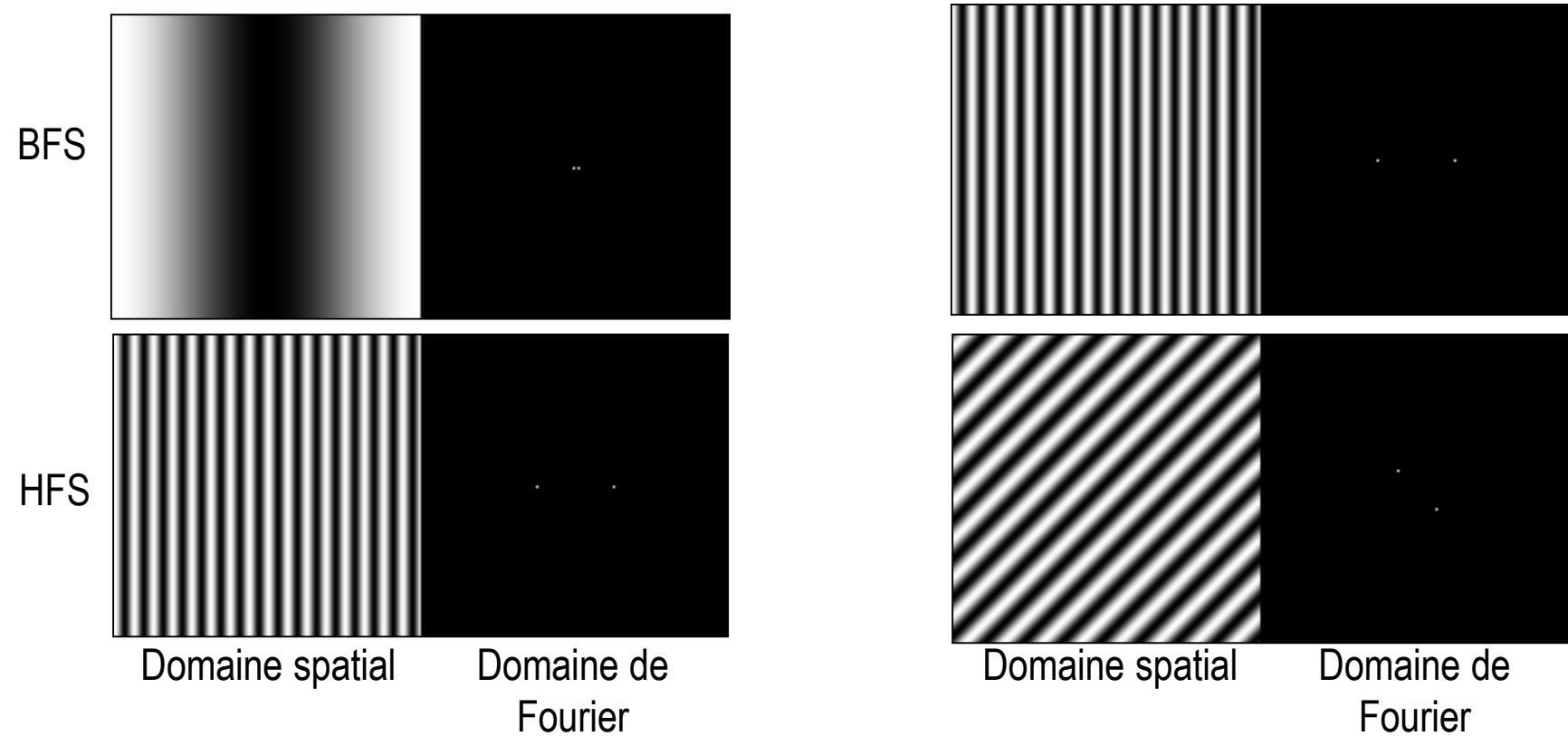


Sinusoïde

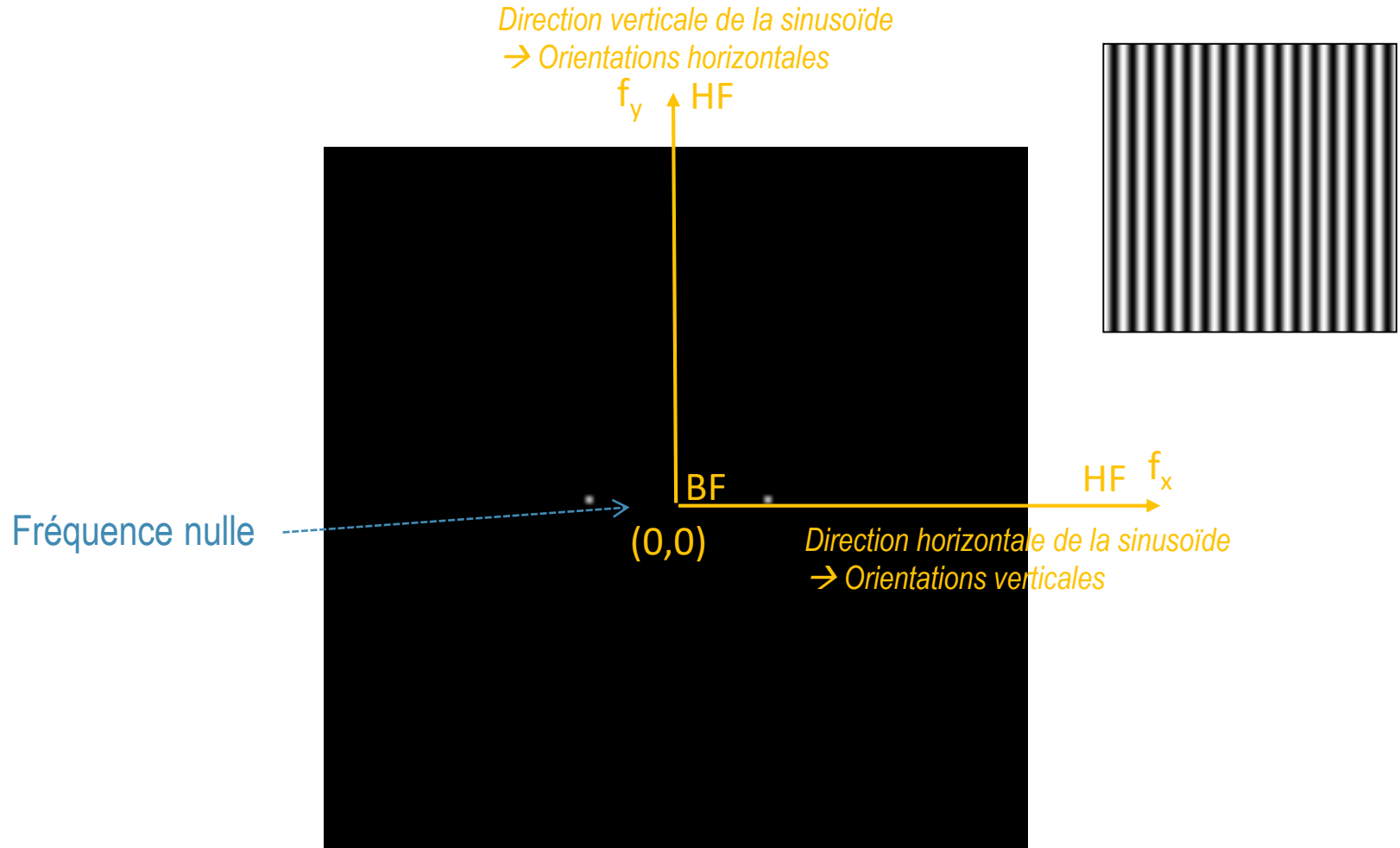
Somme de Sinusoïdes

Application à une image

- Ainsi, la transformée de Fourier d'une image permet d'obtenir son spectre d'amplitude et de phase = répartition de l'amplitude et de la phase de l'énergie de l'image en fonction des fréquences spatiales (FS) ...et des orientations

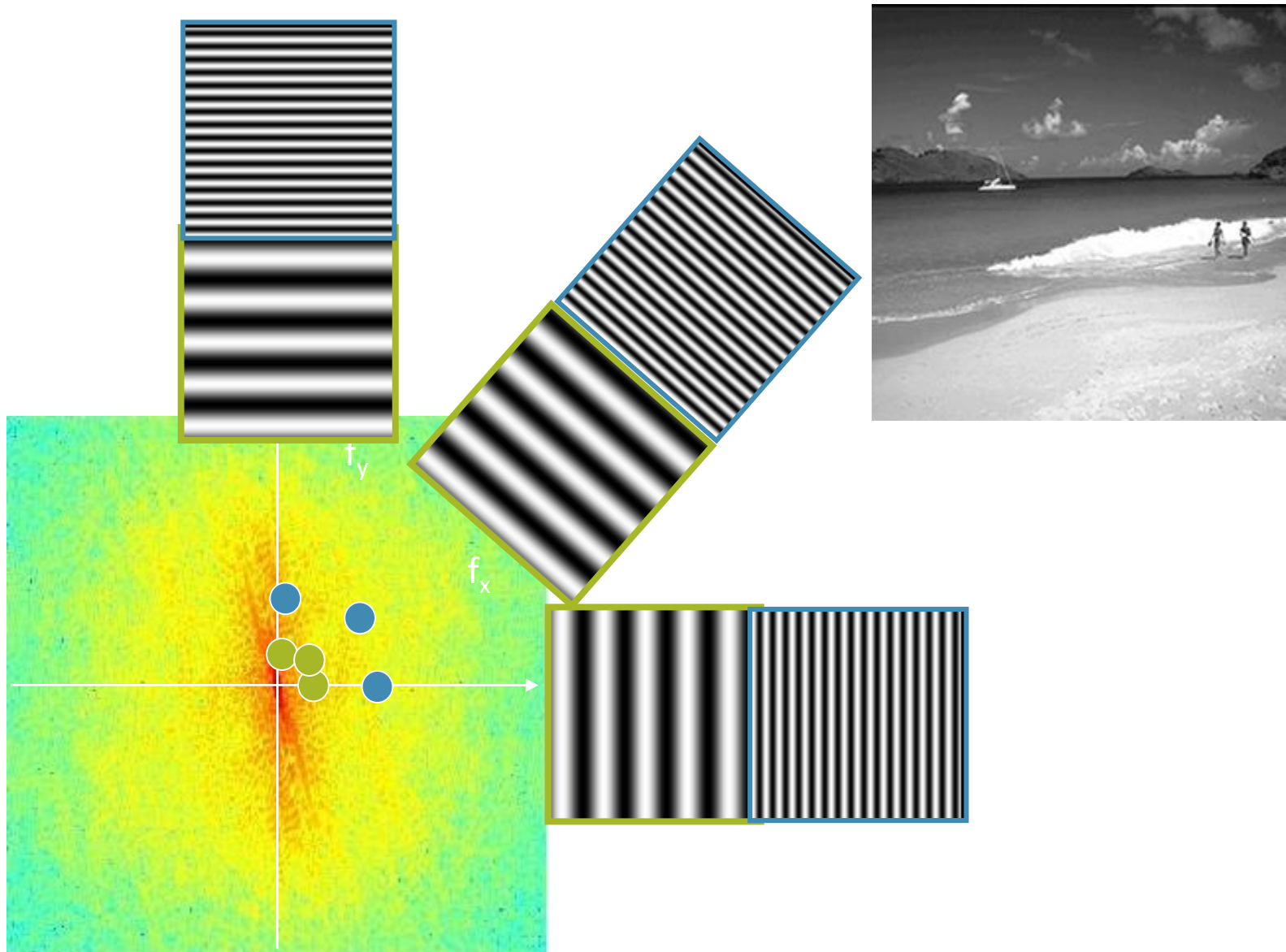


Spectre d'amplitude d'une image



Classiquement, on représente le spectre entre $-f_e/2$ et $+f_e/2$ (f_e étant la fréquence d'échantillonnage)

Spectre d'amplitude d'une scène naturelle

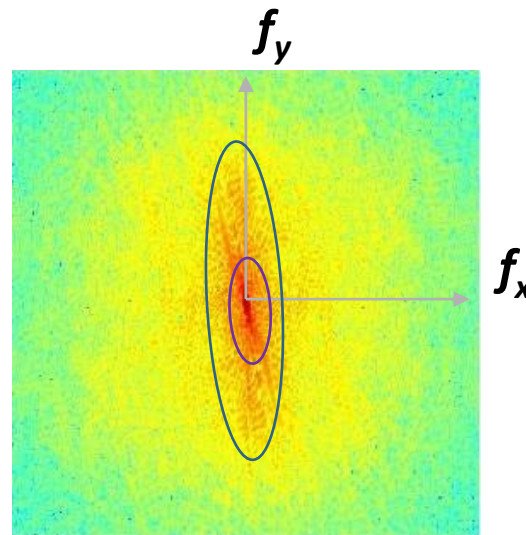


Transformée de Fourier d'une scène naturelle

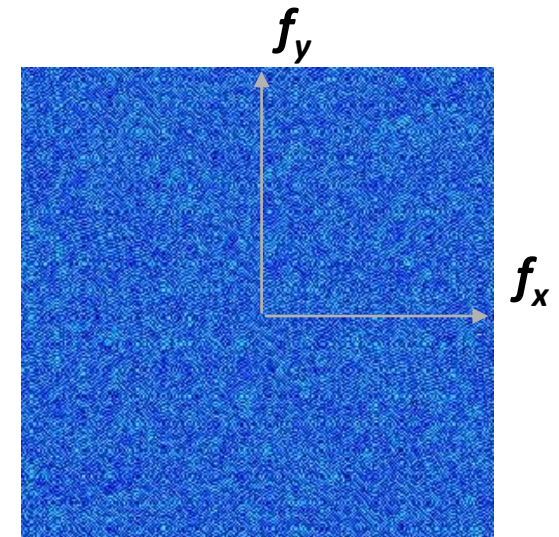


$$I(f_x, f_y) = \sum i(x, y) e^{-j2\pi(f_x x + f_y y)}$$

Spectre d'amplitude (SA)



Spectre de phase (SP)



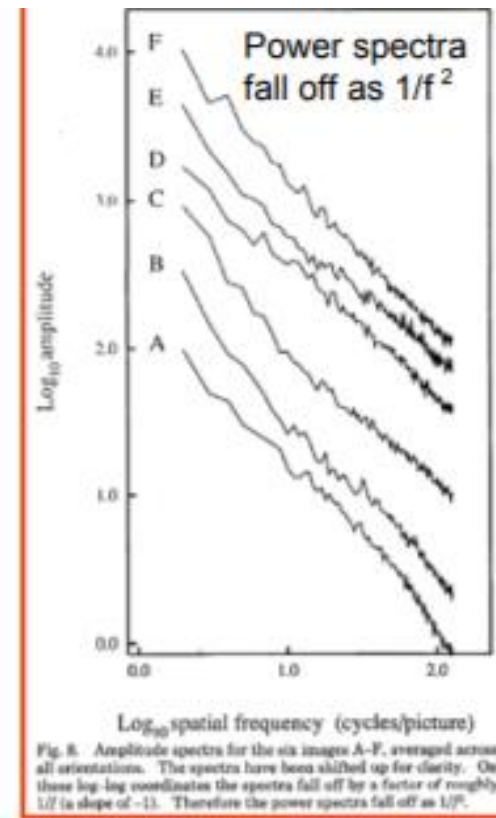
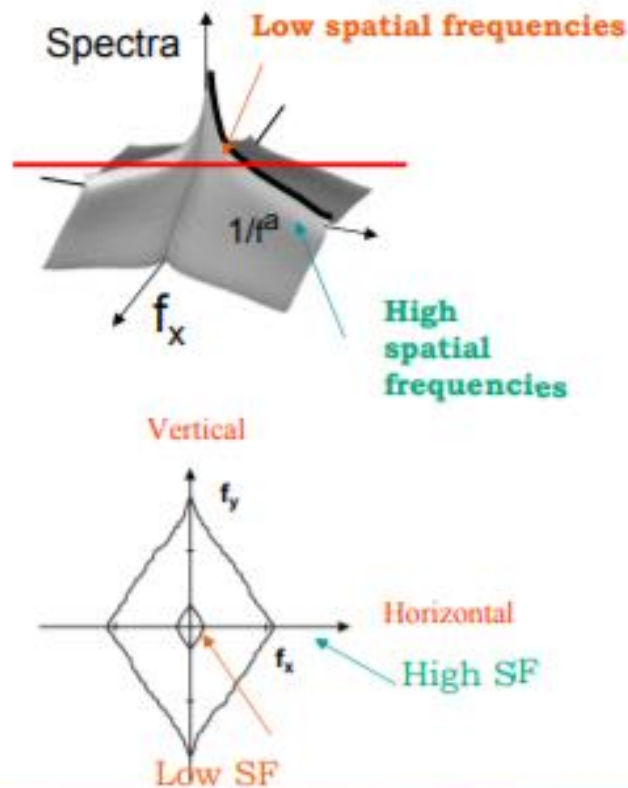
→ Prédominance
d'orientations horizontales

→ Prédominance des basses
fréquences spatiales

$$|I(f_x, f_y)| = \sqrt{\text{Re}[I(f_x, f_y)]^2 + \text{Im}[I(f_x, f_y)]^2} \quad \varphi(u, v) = \text{Arctg} \left(\frac{\text{Im}[I(f_x, f_y)]}{\text{Re}[I(f_x, f_y)]} \right)$$

Régularité statistique des signaux naturels

- Tous signal naturel (image, son...) comporte une décroissance de son énergie avec la fréquence ($1/f$) (i.e. plus d'énergie en BF qu'en HF)

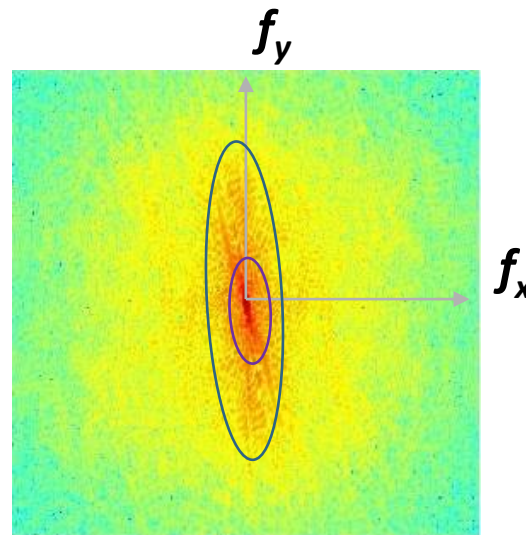


Transformée de Fourier d'une scène naturelle

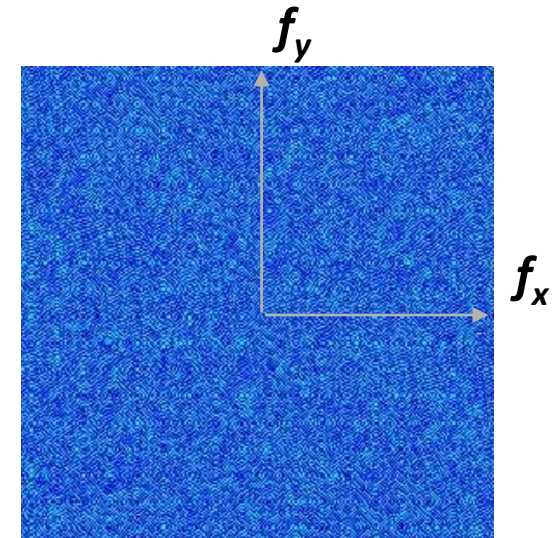


$$I(f_x, f_y) = \sum i(x, y) e^{-j2\pi(f_x x + f_y y)}$$

Spectre d'amplitude (SA)



Spectre de phase (SP)



→ Prédominance
d'orientations horizontales

→ Prédominance des basses
fréquences spatiales

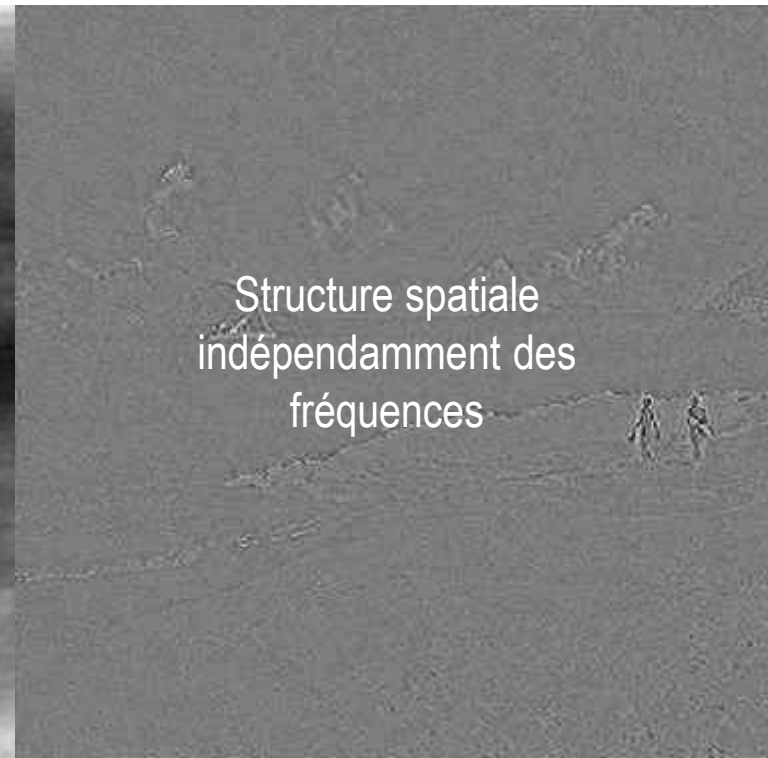
$$|I(f_x, f_y)| = \sqrt{\text{Re}[I(f_x, f_y)]^2 + \text{Im}[I(f_x, f_y)]^2} \quad \varphi(u, v) = \text{Arctg} \left(\frac{\text{Im}[I(f_x, f_y)]}{\text{Re}[I(f_x, f_y)]} \right)$$

Transformée de Fourier d'une scène naturelle

- Quelle information dans le spectre d'amplitude (SA) et dans le spectre de phase (SP)?



SA de l'image de plage avec SP d'un bruit aléatoire



SA d'un bruit avec SP de l'image de plage

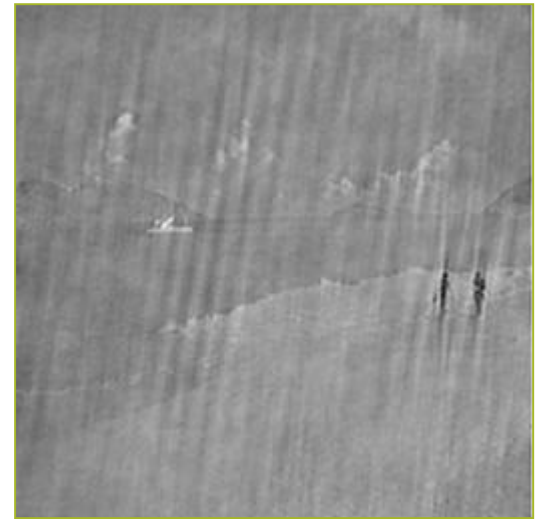
Exemple de manipulation d'images dans le domaine de Fourier: Images « Chimères »



TF

SA

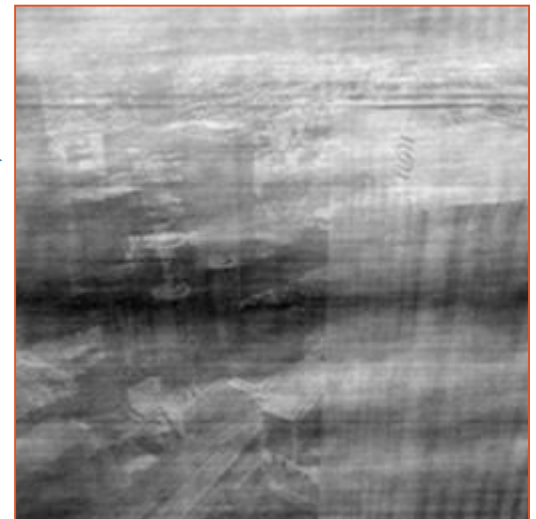
SP

TF
inverse

TF

SA

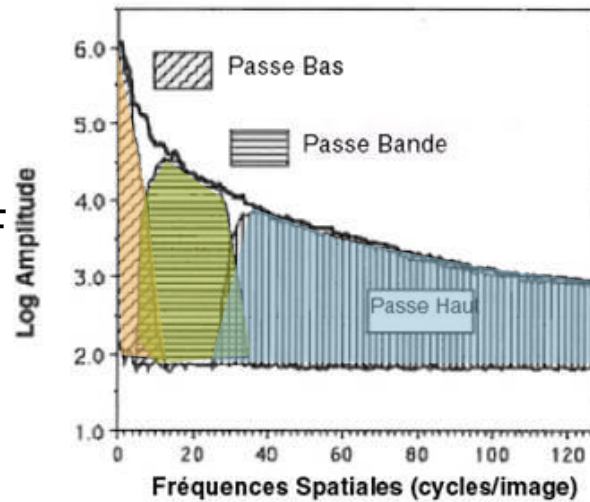
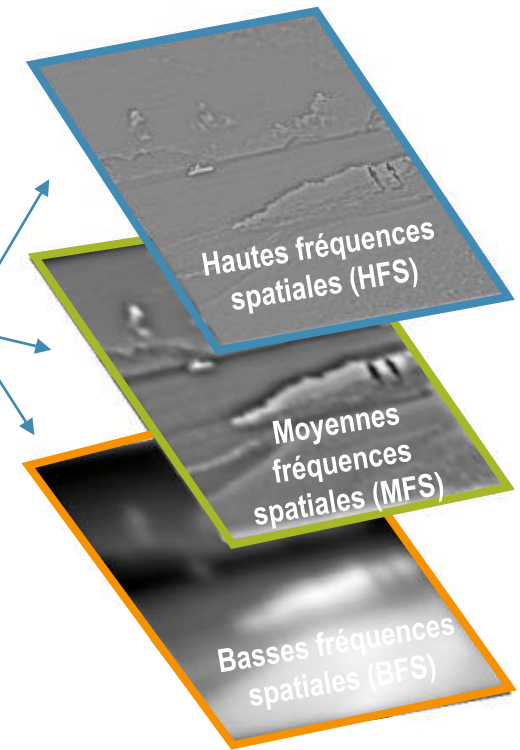
SP

TF
inverse

Exemple de manipulation d'images dans le domaine de Fourier: Filtrage



TF

TF
inverse

Revenons à nos moutons...

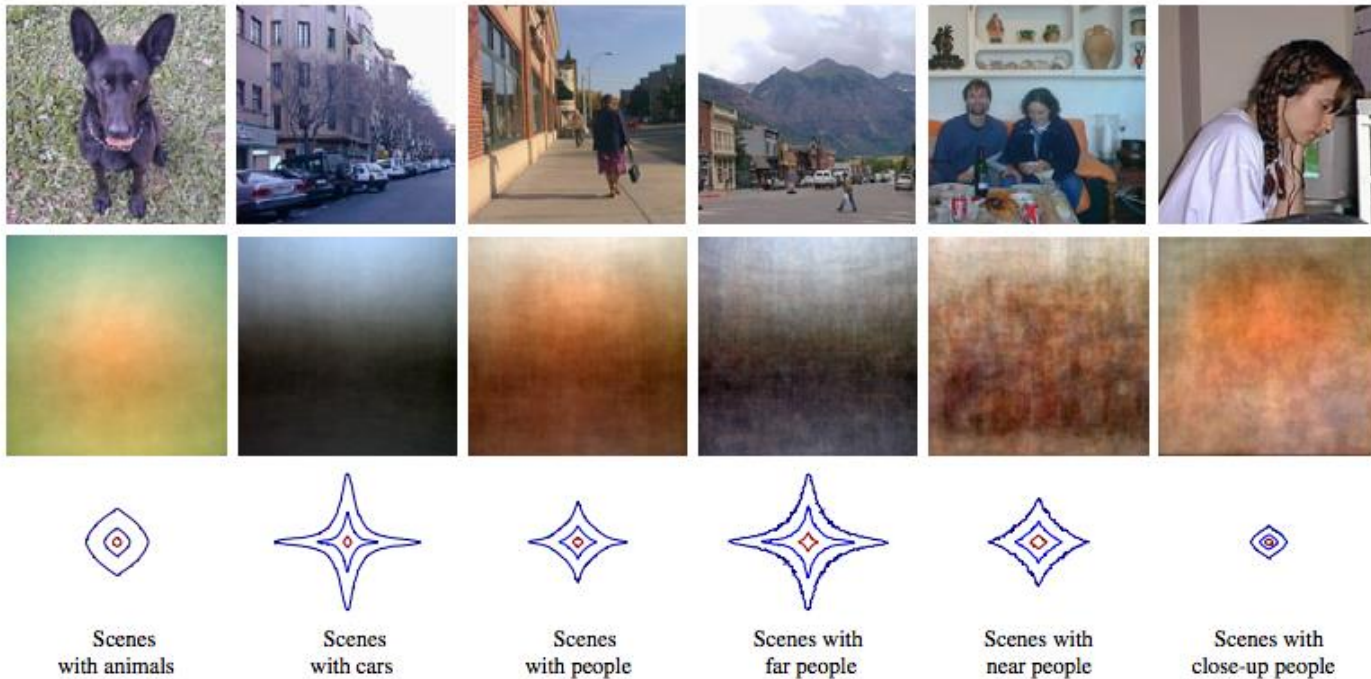
- Comment définir des catégories de scènes?
 - Quelles propriétés sont communes aux exemplaires d'une catégorie?
 - Quelles propriétés sont différentes entre différentes catégories?
- Dans le domaine spatial:
 - Il est difficile de faire émerger des régularités
 - Des calculs de distance ne rendent pas compte des différences perçues



1] Comparison of image fidelity measures for "Einstein" image altered with different types of distortions. (a) Reference image. ... (b) Near contrast stretch. (c) Luminance shift. (d) Gaussian noise contamination. (e) Impulsive noise contamination. (f) JPEG compression. (g) Blurring. (h) Spatial scaling (zooming out). (i) Spatial shift (to the right). (j) Spatial shift (to the left). (k) Rotation (counter-clockwise). (l) Rotation (clockwise).

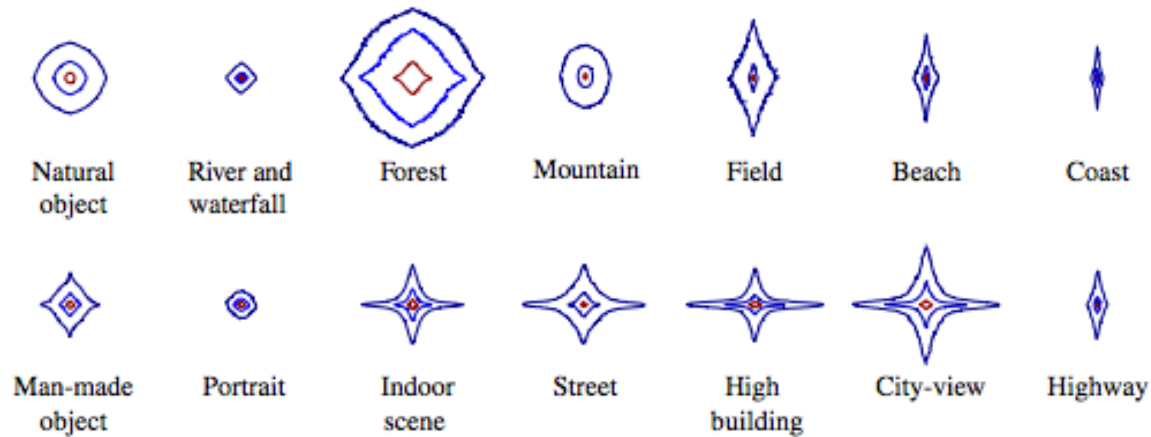
Revenons à nos moutons...

- Comment définir des catégories de scènes?
 - Quelles propriétés sont communes aux exemplaires d'une catégorie?
 - Quelles propriétés sont différentes entre différentes catégories?
- Dans le domaine de Fourier? → Le SA moyen semble être représentatif de catégories de scènes



Descripteurs de catégories de scènes

- Le SA semble être un bon descripteur de la catégorie des scènes



- Une scène de scènes?

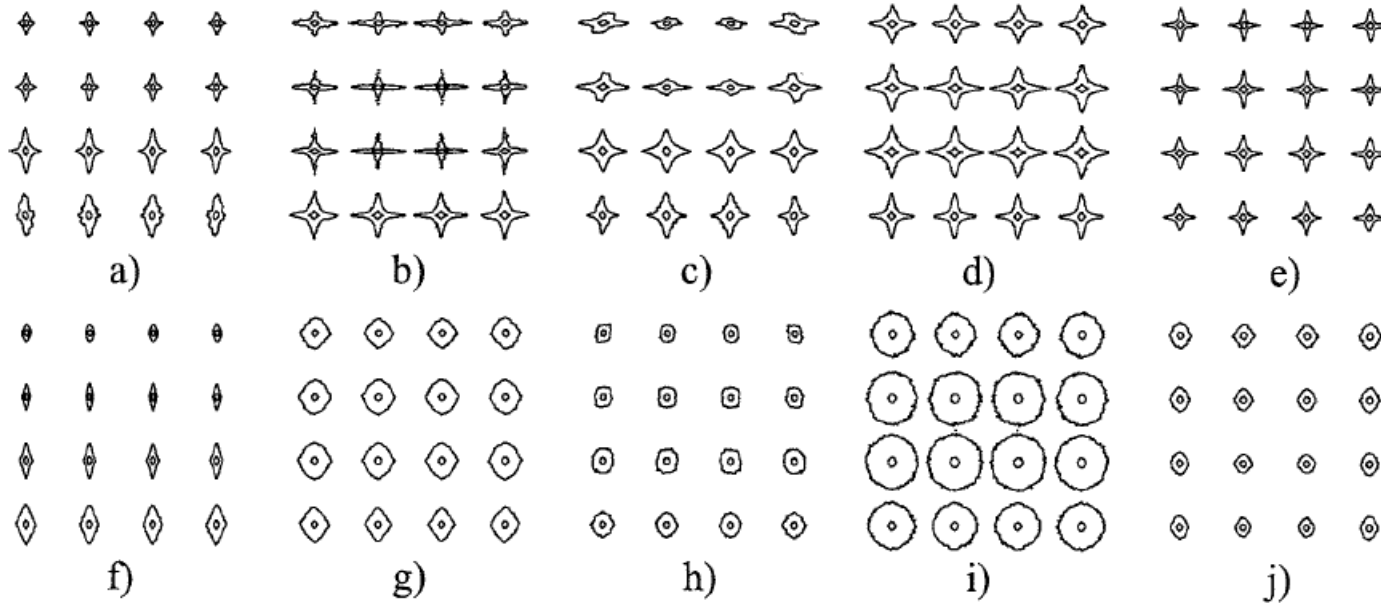


➔ Faire des TF locales



Transformées de Fourier Locales

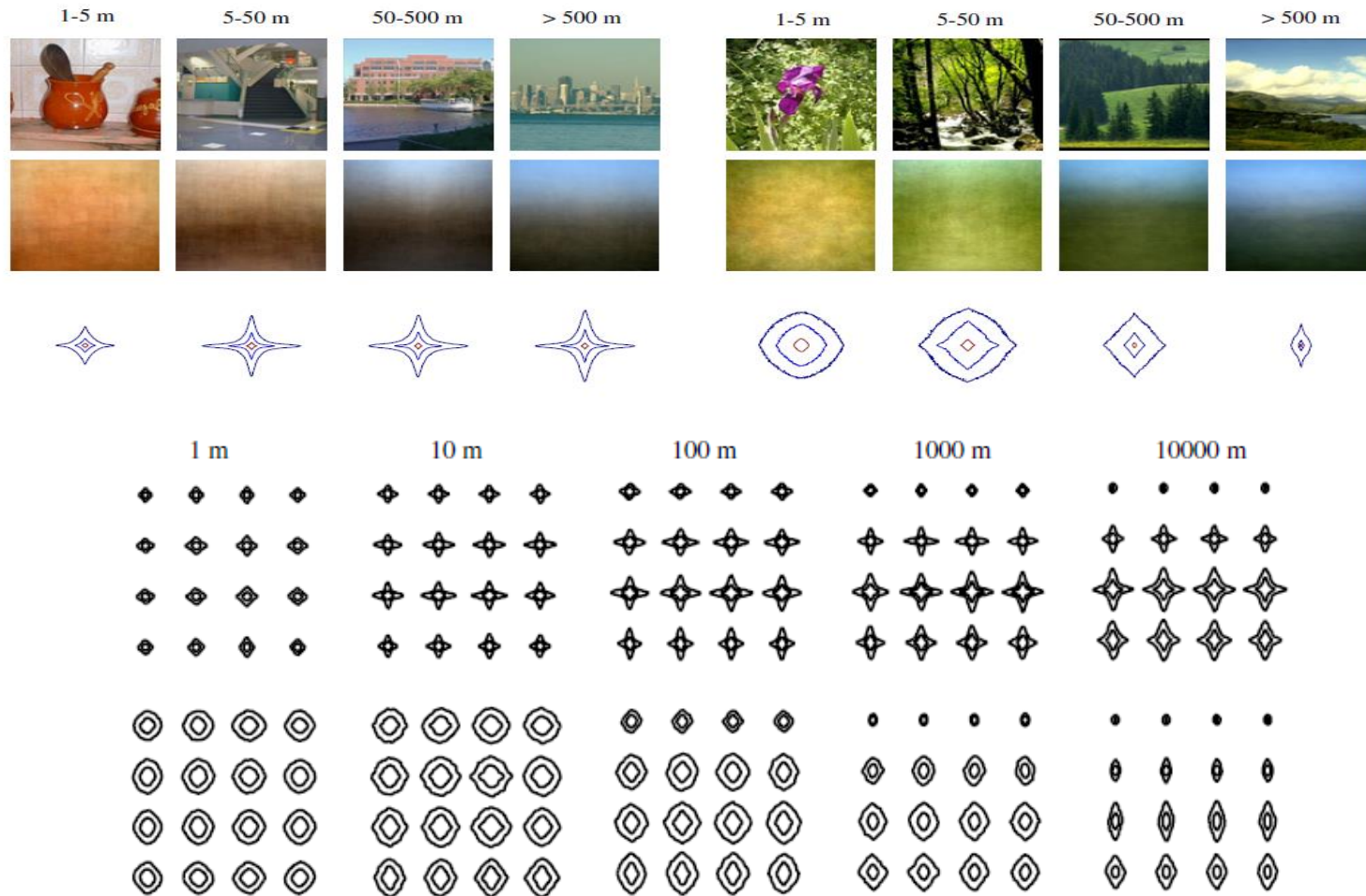
- L'image est découpée en 16 imagettes (4 x 4)
 - On peut ainsi voir l'évolution spatiale des statistiques



man-made open (a) and urban vertically structured (b) environments, perspective views of streets (c), far view of city-center buildings (d) and close-up views of outdoor urban structures (e) and natural open (f) and enclosed (g) environments, mountainous landscapes (h), enclosed forests (i) and close-up views of non-textured natural structures like rocks and water (j). It must be

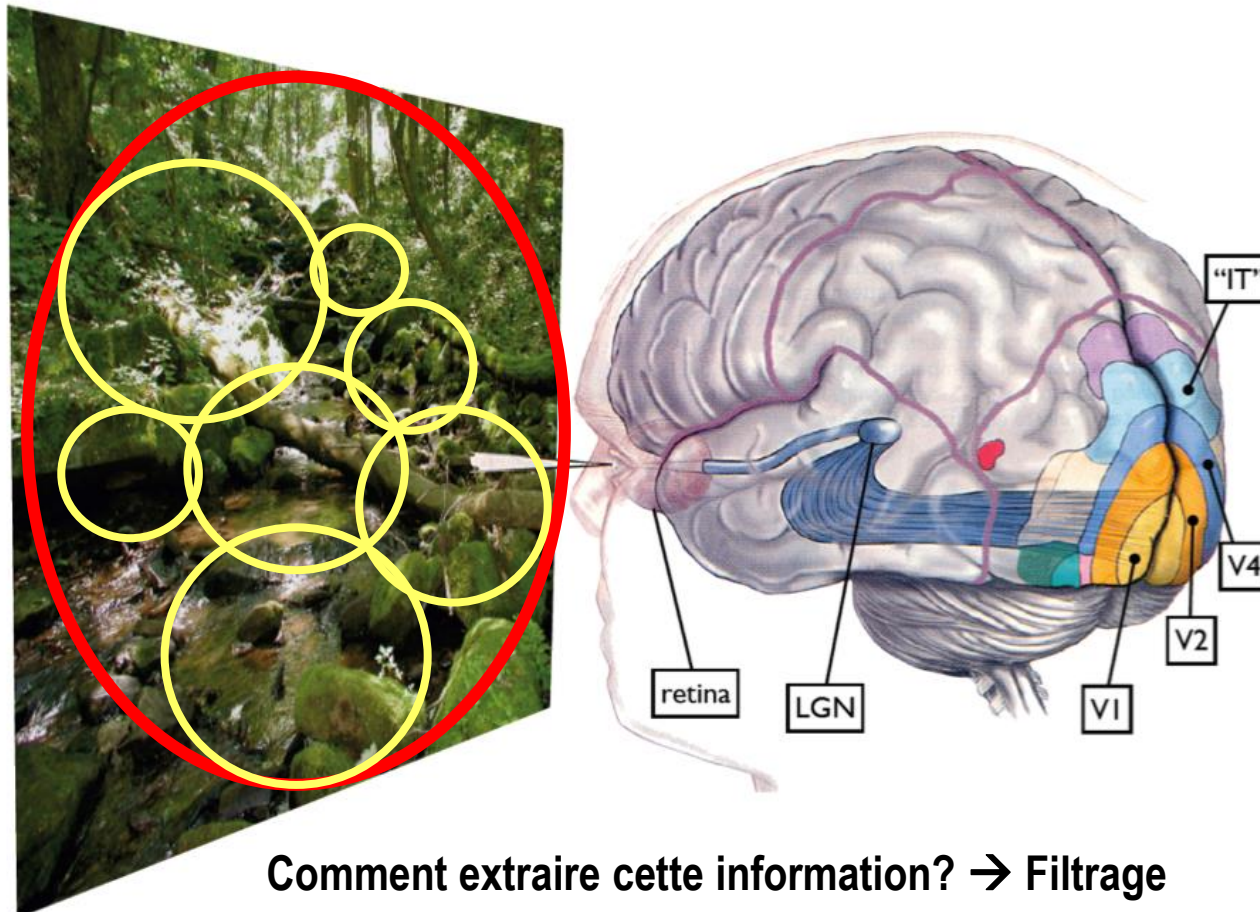
Transformées de Fourier Locales

- En plus de la catégorie, les TF locales permettent de rendre compte d'autres descripteurs des scènes comme la profondeur de champ



Lien avec le système visuel?

- Les premières étapes de l'analyse visuelle peuvent être comparées à une analyse de la scène dans le domaine de Fourier: extraction en parallèle d'une information **localisée** à différentes **orientations** et différentes **fréquences spatiales**



Comment extraire cette information? → Filtrage