

*République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie Houari Boumediene
Faculté d'Electronique et d' Informatique
Département d'Informatique*

Traitement et Analyse d'images et de la vidéo

Master Informatique Visuelle

Chapitre 2

Prétraitement de l'Image et Extraction de Primitives

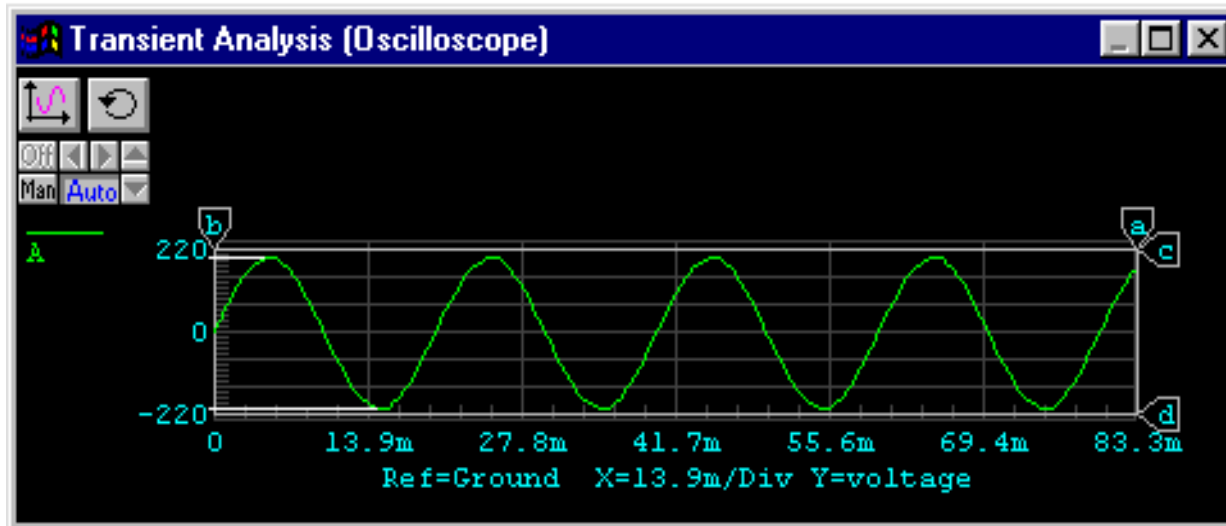
Cours de Traitement et Analyse d'images et de la vidéo
Master MIV, TAI- par N.BAHA

2.2 Filtrage dans le domaine fréquentiel

- le traitement du signal est largement utilisé en Télécommunication dans des applications diverses et variées comme les Télécommunications tel que la protection d'information contre le bruit.
- Le but est la caractérisation d'un signal dans le domaine temporel et fréquentiel pour aboutir à des modèles mathématiques. La description mathématique des signaux permet de concevoir et de caractériser des systèmes de traitement de l'information. Le bruit représentera un « signal » ou phénomène perturbateur

Notions de Base du traitement du signal

Habituellement, on représente un signal par une fonction continue dans le temps et on le visualise sur un appareil représentant la variation d'amplitude d'un phénomène en fonction du temps : cardiogramme, sismographe...

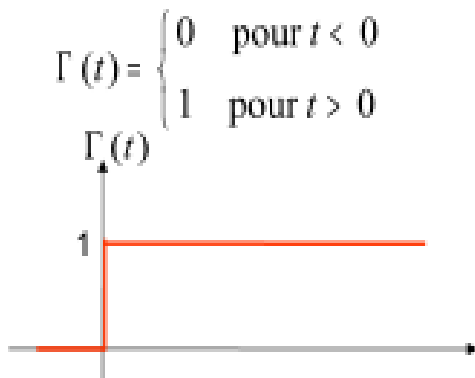


Notions de Base du traitement du signal

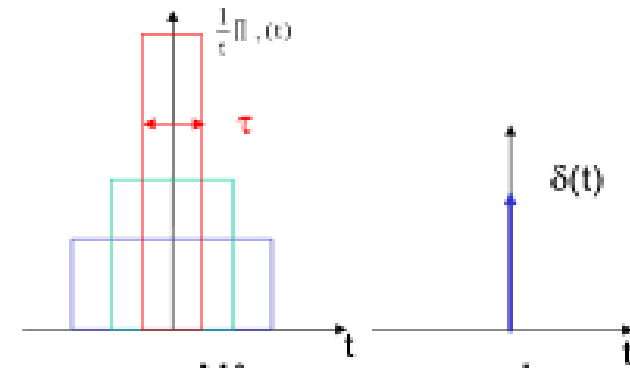
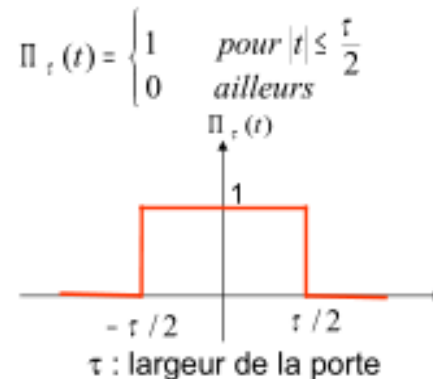
On définit deux classes principales de signaux :

- **Signal déterministe** : il s'agit d'un signal dont on peut représenter l'évolution grâce à une **fonction mathématique**. On peut citer le signal sinusoïdal, rampe, échelon, impulsion ou dirac, ...

■ Échelon



■ Signal porte ou rectangle



Notions de Base du traitement du signal

- Un signal déterministe peut être périodique ou non périodique.
- **Signal aléatoire** est un signal dont on ne peut deviner l'évolution. Néanmoins, tout signal aléatoire peut être caractérisé mathématiquement, mais aucune fonction mathématique ne permet de prédire l'évolution du signal à l'instant donné. Un signal aléatoire peut être stationnaire ou non stationnaire.

Notions de Base du traitement du signal

Définition :

- Un signal déterministe, représenté par sa fonction u est dit périodique de période T_p si

$$u(t)=u(t+T_p).$$

- Propriétés :

Tout signal périodique de période T_p présente une fréquence fondamentale à la fréquence $f_p=1/T_p$.

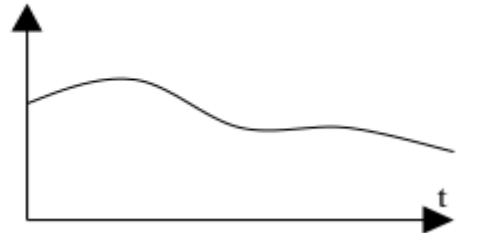
- La Moyenne : est définie par l'intégrale de la fonction sur une période (d'où la nécessité d'un signal périodique).

$$m = \frac{1}{T} \int_0^T u(t) dt$$

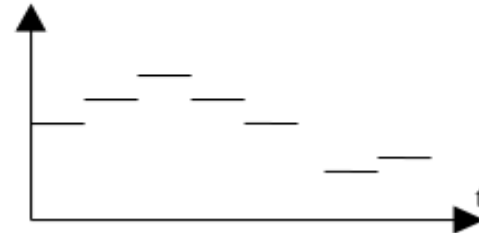
Notions de Base du traitement du signal

- La puissance P ou valeur efficace
Seff:est définie par l'intégrale de la
fonction au carré

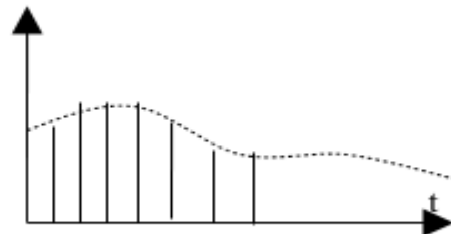
$$S_{eff}^2 = P = \frac{1}{T} \int_0^T u^2(t) dt$$



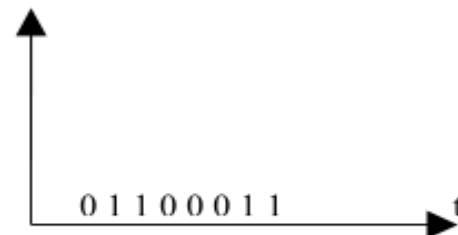
Signal continu dans le temps



Signal discretisé



Signal échantillonné



Signal Numérique

Notions de Base du traitement du signal

En résumé il existe différents types de signaux: déterministes et aléatoires. Les premiers sont connus par l'utilisateur, les seconds sont dus aux bruits ou sont destinés à modéliser des systèmes. Ces signaux sont couramment utilisés en traitement du signal, on les associe à leur puissance, et aux propriétés de corrélations. Nous allons couramment mesurer leur spectre, et l'évolution de leur spectre.

Notions de Base du traitement du signal

Pourquoi filtrer un signal?

Filtrage : élimination de certaines composantes par exemple:

- Détection de craquements sur un enregistrement,
- Détection de bruit sur une image,
- Annulation d'écho, etc.

La suite du cours est destiné à l'étude des séries et des **Transformées de Fourier(TF)** qui permet l'analyse fréquentielle d'un signal. On parle de :

▲ TF continue, TF discrète, FFT(Fast Fourier Transform)

Transformation d'image

Série De Fourier

Tout signal $x(t)$ périodique peut être écrit sous un développement en séries de Fourier à partir de signaux de base de formes exponentielle complexe.

Avec

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega_p t) + b_n \sin(n\omega_p t)$$

$$a_0 = \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt, \text{ est la composante continue}$$

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos(n\omega_p t) dt, \quad b_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin(n\omega_p t) dt, \quad n > 1$$

Les coefficients a_n et b_n sont appelés les coefficients de la série de Fourier,
 $\omega_p = 2\pi f_p$ est la pulsation avec $f_p = 1/T$.

Transformation d'image

La Transformée de Fourier:

C'est un outil fondamental d'analyse en traitement du signal.

La transformée de Fourier permet la décomposition d'un signal f en combinaison linéaire de sinusoides complexes, dont les coefficients $F[u,v]$ dit coefficients de Fourier, fournissent des informations sur les fréquences (u,v) et permettent des manipulations dans le domaine fréquentiel.

Application à une image

Soit x une fonction à deux variables représentant l'intensité d'une image. La transformée de Fourier de cette image permet de passer d'une représentation spatiale à la représentation de l'image dans le domaine fréquentiel.

Transformation d'image

Application à une image

Basses fréquences →

- Lentes et faibles variations
- zones presque uniformes

Hautes fréquences →

- Variations rapides et contrastées
- Contours
- Détails
- Bruit



Transformée de Fourier

Transformée de Fourier **discrète** bidimensionnelle d'une Image de dimension (N,M): largeur N et de hauteur M est:

x et y sont les coordonnées
du domaine spatial

Directe

$$F(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) e^{-i2\pi(\frac{ux}{N} + \frac{vy}{M})}$$

u et v sont les coordonnées
du domaine fréquentiel

Inverse

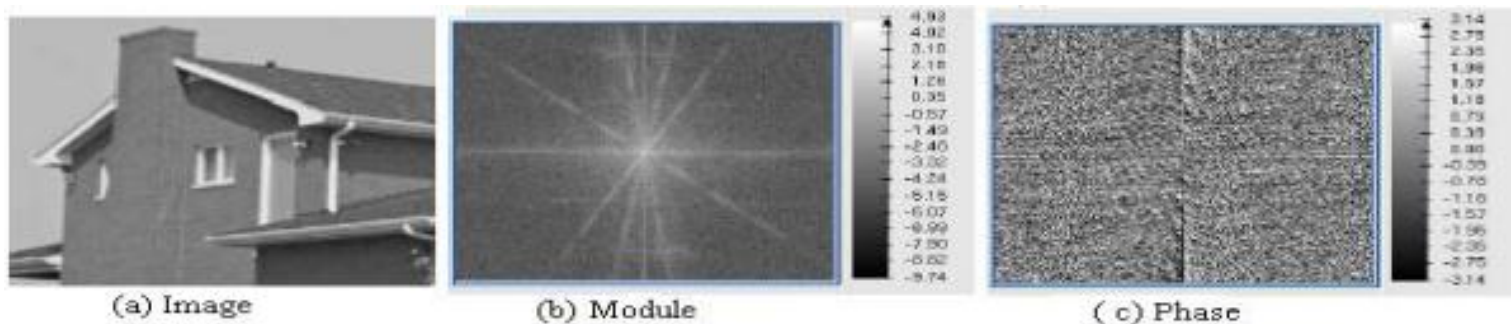
$$f(x, y) = \frac{1}{MN} \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} F(u, v) e^{i2\pi(\frac{ux}{N} + \frac{vy}{M})}$$

On l'appelle la transformée de Fourier Inverse.

Transformée de Fourier

La Transformée de Fourier se représente dans un espace fréquentiel bidimensionnel. Etant donné que la transformée est une grandeur **complexe**, sa représentation **graphique** se fait soit par

- le module et la phase,
- la partie réelle et imaginaire de la transformée de Fourier.



Transformée de Fourier

Transformée de Fourier **discrète du signal**

(x_0, x_1, \dots, x_n) correspond à (X_0, X_1, \dots, X_n) donné par la :

$$X_n = \sum_{k=0}^{N-1} x_k e^{\frac{-i2\pi nk}{N}}$$

Transformée de Fourier **discrète inverse**:

$$x_n = \sum_{k=0}^{N-1} X_k e^{\frac{+i2\pi nk}{N}}$$

Transformée de Fourier

Exemple

Calculer la TFD de la séquence (0,1,0)

$x_0=0, x_1=1, x_2=0$ donc $N=3$

$$X_n = \sum_{k=0}^{N-1} x_k e^{\frac{-i2\pi nk}{N}}$$

Donc nous allons calculer la Transformée de Fourier **discrète du signal**

(x_0, x_1, \dots, x_n) correspond à (X_0, X_1, \dots, X_n) donné par la :

$$X_0 = \sum_{k=0}^2 x_k e^{\frac{-i2\pi 0k}{N}} = x_0 e^{\frac{-i2\pi 00}{3}} + x_1 e^{\frac{-i2\pi 01}{3}} + x_2 e^{\frac{-i2\pi 02}{3}} = 0 + 1 + 0 = 1$$

$$\begin{aligned} X_1 &= \sum_{k=0}^2 x_k e^{\frac{-i2\pi 1k}{N}} = x_0 e^{\frac{-i2\pi 10}{3}} + x_1 e^{\frac{-i2\pi 11}{3}} + x_2 e^{\frac{-i2\pi 12}{3}} \\ &= x_0 + x_1 e^{\frac{-i2\pi}{3}} + x_2 e^{\frac{-i4\pi}{3}} = 0 + e^{\frac{-i2\pi}{3}} + 0 = e^{\frac{-i2\pi}{3}} \end{aligned}$$

$$\begin{aligned} X_2 &= \sum_{k=0}^2 x_k e^{\frac{-i2\pi 2k}{N}} = x_0 e^{\frac{-i2\pi 20}{3}} + x_1 e^{\frac{-i2\pi 21}{3}} + x_2 e^{\frac{-i2\pi 22}{3}} \\ &= e^{\frac{-i4\pi}{3}} \end{aligned}$$

Transformée de Fourier

Exemple

$$= 1 + e^{\frac{-i2\pi}{3}} + e^{\frac{-i4\pi}{3}}$$

Remarque

Le calcul de la TFD engendre beaucoup de calculs d'où la **FFT** (Fast Fourier Transform).

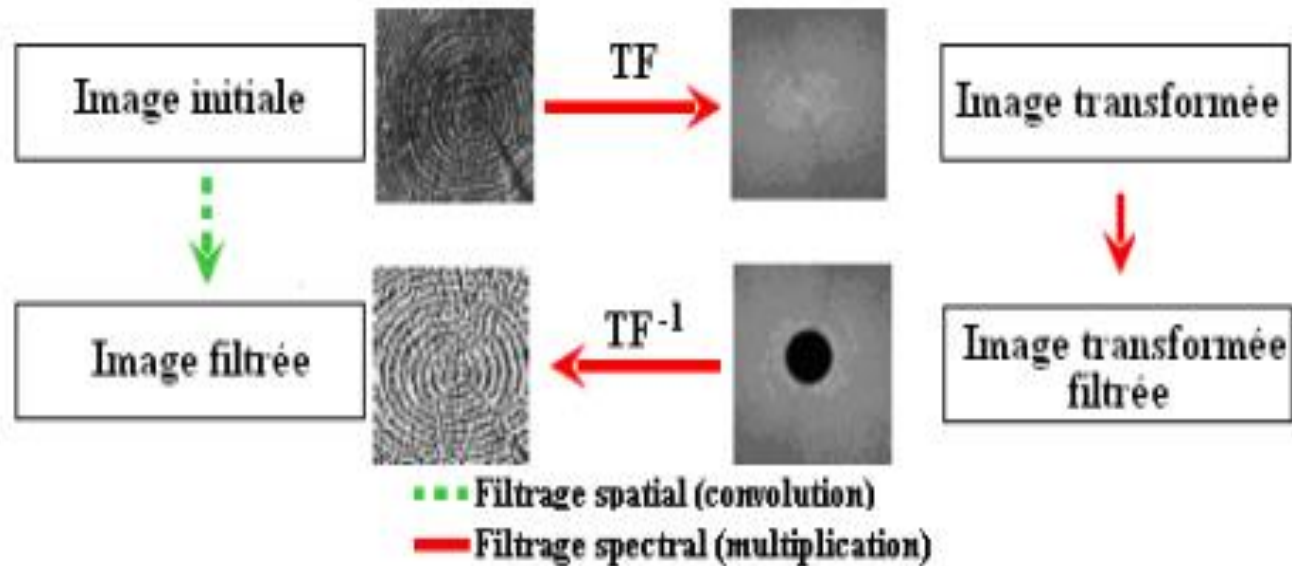
Transformée de Fourier Discrète

- Il est possible d'extraire la composante fréquentielle de l'image avec des filtres passe haut, passe-bas ou passe bande.
- Un filtrage **passe-haut** laisse passer les **hautes fréquences** et atténue les basses fréquences (il a pour effet de faire **apparaître les détails de l'image**).
- Un filtrage **passe-bas** laisse passer les **basses fréquences** et atténue les hautes fréquences (il a pour effet de faire **disparaître les détails de l'image**).
- Etant donné la propriété de linéarité de la transformée de Fourier, l'image peut être décomposée comme

$$\text{IMG} = \text{BF} + \text{HF}.$$

Transformée de Fourier Discrète

- Résumé

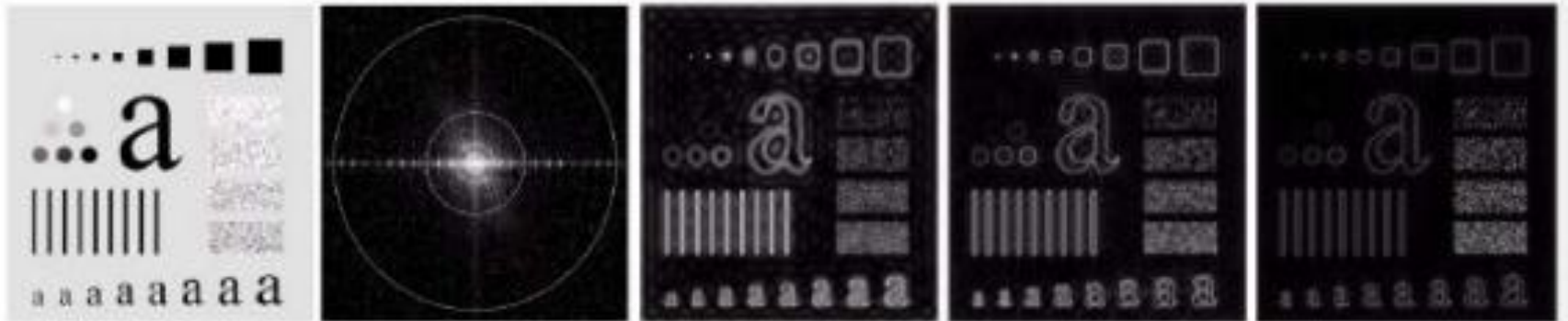


- Remarque:** une multiplication dans le domaine des fréquences (spectral) est équivalente à une convolution dans le domaine spatial, c'est-à-d

$$f(t) * s(t) \Leftrightarrow F(\omega)S(\omega):$$

Transformée de Fourier Discrète

- Exemple



Image, son spectre et les trois images obtenues par des filtres passe haut de rayon 15,30 et 80

Transformée de Fourier Discrète

Principe de filtrage dans le domaine fréquentiel

Lisser un signal ou une image revient à éliminer les composantes fréquentielles dont les valeurs sont faibles ou hautes selon le type du bruit.

- 1- Calculer la TF de l'image,
- 2- Appliquer le filtre (éliminer les fréquences correspondantes aux bruits).
- 3- Restituer l'image en calculant la transformée de Fourier Inverse du spectre obtenu après avoir filtrer l'image .

Extraction de primitives

- Les différentes phases de l'analyse d'une image

- 1- Acquisition
- 2- Traitement de bas niveaux : filtrage et extraction de primitives.
- 3- Traitement de haut niveaux : reconnaissance des formes, reconstruction, calculs de mouvements, interprétation d'une scène.....

- Les éléments impliqués

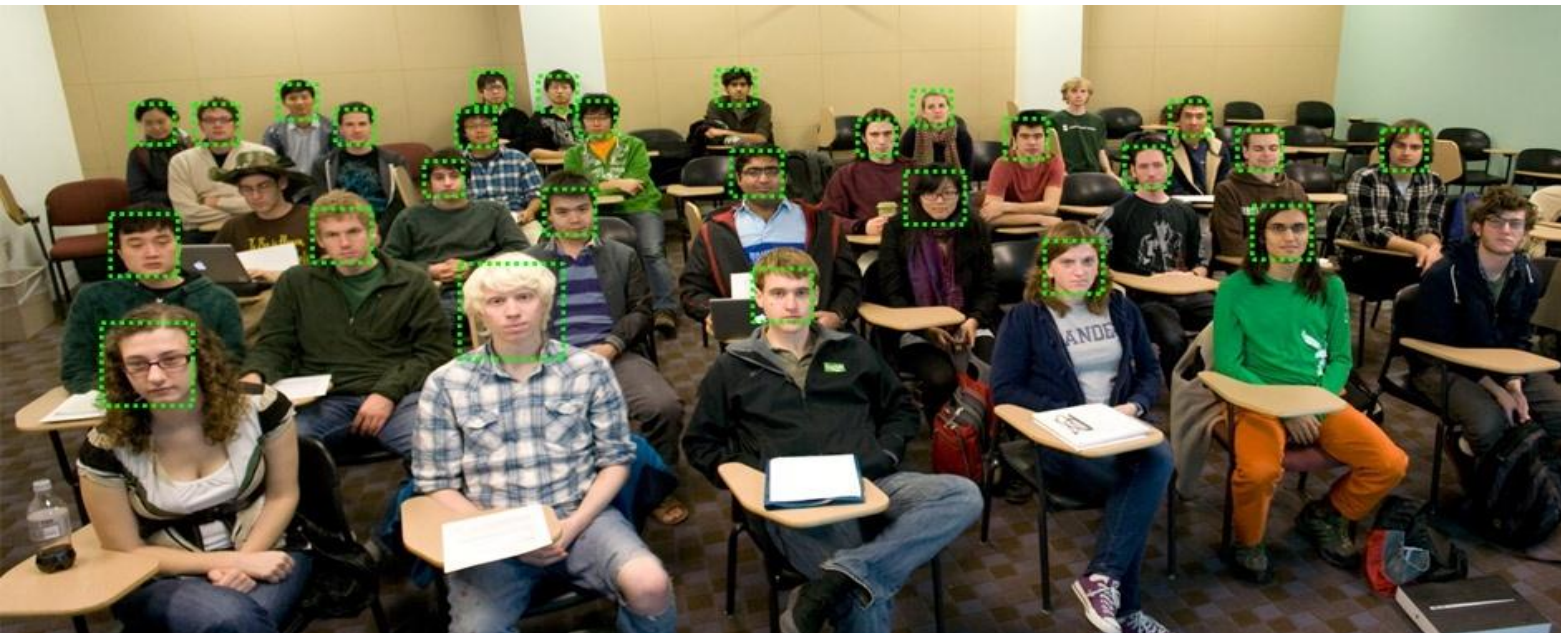
- 1- **Des primitives images** : pixels, points d'intérêts, segments, contours.
- 2- **Des caractéristiques photo-métriques** : niveaux de gris, de couleurs, Texture.
- 3- **Des caractéristiques géométriques** : caméras, Kinect, mouvements.

Extraction de primitives

Détection des points d'intérêts



Détection de visages



Extraction des primitives

Qu'est ce qu'une primitive dans l'image ?

- Localiser, mesurer, ou reconnaître un objet vu dans une image exige de trouver quels sont les pixels de l'image correspondants à cet objet.
- L'extraction de différentes primitives permet d'obtenir une description compacte de l'image qui est plus exploitable.

■ Définition

Une primitive est l'élément de base dans une image à partir duquel on peut construire des images.

- Ces primitives peuvent être soit
 - Des pixels caractéristiques (pixels d'intérêts),
 - Des contours,
 - Des régions.

Extraction des primitives

Il existe une dualité entre ces différentes primitives.

- Une région définit une chaîne de contours fermés qui à son tour définit un ensemble de pixels caractéristiques. Inversement, un ensemble de pixels définissant une chaîne de contours fermée, définit une région.
- *L'extraction des primitives est une étape importante qui constitue ce que l'on appelle le bas niveau dans le système de la vision artificielle.*
- Le but de cette étape est d'extraire d'une image, un ensemble de données définies et structurées de telle sorte à pouvoir catégoriser une image.

Extraction des primitives

- Pourquoi détecter des primitives ?
 - Pour les utiliser dans l'analyse d'image.
- Objectifs
 - Reconstruction 3D
 - Analyse de scènes,
 - Interprétation,
 - Détection d'objets....
- Deux étapes importantes :
 - 1- Détection de la primitive (feature)
 - 2- Description, caractérisation de la primitive.

Détection des Points d'intérêts

Définition des Points d'Intérêts: Pol

- L'idée derrière cette appellation est que lorsqu'un être humain regarde une image, il lui suffit de regarder certains points ou régions d'intérêts pour identifier ou reconnaître les objets présents dans l'image.
- Donc l'être humain est capable d'identifier rapidement des caractéristiques visuelles importantes de l'image grâce à ces points qui contiennent beaucoup d'informations .
- Les points d'intérêts extraits à partir d'une image permettent de caractériser cette image. En comparant les points d'intérêts d'une image et ceux d'une autre image, on peut alors déduire si des informations communes sont présentes dans ces deux images.

Détection des Points d'intérêts

Détection des pixels d'intérêts: features

- La détection de pixels d'intérêts (ou *coins*) est au même titre que la détection de contours, une étape préliminaire à de nombreuses applications de la vision par ordinateur.
- De nombreuses méthodes ont été proposées pour détecter les pixels d'intérêts:
 - Approches contours,
 - Approches intensité.

1- Approches contours : On commence par détecter les contours dans une image dans un premier temps. Les points d'intérêts sont ensuite extraits le long des contours en considérant les points de courbures maximales ainsi que les intersections de contours.

2. Approches intensité : consiste à examiner directement la fonction d'intensité dans l' image pour en extraire directement les points de discontinuités.

Détection des Points d'intérêts

- Les approches de la deuxième catégorie sont celles utilisées car elles sont : indépendantes vis à vis de la détection de contours (stabilité).
- De nombreux détecteurs de pixels existent:
 - Détecteur de Moravec [MOR80];
 - Détecteur Hessian Beaudet78
 - Détecteur d'Harris et Stephens, [1988];
 - Détecteur de Smith et Brady, [1997].
 - SIFT
 - SURF

Détection des Points d'intérêts

Point d'intérêts ou Features POI

➤ Propriétés attendues

- **Robustesse**: Invariance aux changements de :
luminosité, d'échelle, rotation, transformations affines, perspective.

Le point doit apparaître aux mêmes endroits quelque soit la déformation.

- **Distinctif** : utile pour les recherches dans de grandes bases.
- **Efficace** : facile et rapide à calculer (temps-réel).
- **Représentatif** : beaucoup de points/régions détectés même sur des petites zones/objets.

Détection des Point d'intérêts

Un peu d'histoire!

- L'idée d'utiliser des points d'intérêt locaux remonte aux travaux de Hans Moravec en 1981 sur la recherche de correspondance entre images stéréoscopiques et aux améliorations apportées en 1988 par Harris et Stephens.
- En 1992, l'intérêt de ce type de détecteur est confirmé par les travaux de Harris (Harris amélioré) et le descripteur de coins qu'il propose, améliorant la plupart des défauts du détecteur de Moravec, va connaître un important succès. Il apporte l'invariance à la rotation, il reste cependant sensible aux changements d'échelle, d'angle d'observation et changements de luminosité . Lowe comblera grandement ces défauts avec son descripteur **SIFT**.

Détection des Point d'intérêts

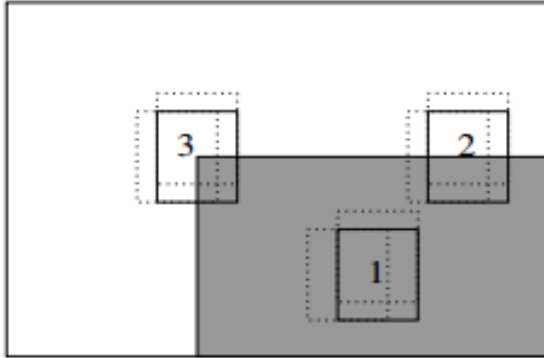
- le détecteur de Moravec (1980)

- L'idée du détecteur de Moravec est de considérer le voisinage d'un pixel (une fenêtre) et de déterminer les changements moyens de l'intensité dans le voisinage considéré lorsque la fenêtre se déplace dans diverses directions. Plus précisément. on considère la fonction :

$$E(x, y) = \sum_{u, v} w(u, v) |I(x + u, y + u) - I(u, v)|^2,$$

- w spécifie la fenêtre/voisinage considérée (valeur 1 à l'intérieur de la fenêtre et 0 à l'extérieur);
- $I(u, v)$ est l'intensité au pixel (u, v) ;
- $E(x, y)$ représente la moyenne du changement d'intensité lorsque la fenêtre est déplacée de (x, y) .

Détection des Point d'intérêts



Les différentes situations traitées par le détecteur de Moravec.

En appliquant cette fonction dans les 3 situations précédentes:

- 1- L'intensité est approximativement constante dans la zone image considérée : la fonction E prendra alors de faibles valeurs dans toutes les directions (x,y) .
- 2- La zone image considérée contient un contour rectiligne : la fonction E prendra alors de faibles valeurs pour des déplacements (x,y) le long du contour et de fortes valeurs pour des déplacements perpendiculaires au contour.
- 3- La zone image considérée contient un coin ou un point isolé : la fonction E prendra de fortes valeurs dans toutes les directions

Détection des Points d'intérêts

- En 1988, Harris et Stephen ont identifié certaines limitations et, en les corrigeant ont déduit un détecteur de coins très populaire : le détecteur de Harris. Ils ont remarqué que Moravec utilisait la Sum of Squared Difference (SSD) pour détecter les points interets ce qui est équivalent à utiliser une matrice d'autocorrélation.

- **Détecteur d'Harris**

Le détecteur de coin Harris est basé principalement sur le fait qu'au niveau d'un coin, l'intensité change grandement dans différentes directions. Le détecteur d'Harris a apporté des améliorations au détecteur de Moravec.

Ils utilisent le développement limité de Taylor de la fonction d'intensité $I(u,v)$ au voisinage du pixel (u,v) :

Détection des Points d'intérêts

$$I(x + u, y + v) = I(u, v) + x \frac{\delta I}{\delta x} + y \frac{\delta I}{\delta y} + o(x^2, y^2).$$

D'où :

$$E(x, y) = \sum_{u, v} w(u, v) \left[x \frac{\delta I}{\delta x} + y \frac{\delta I}{\delta y} + o(x^2, y^2) \right]^2,$$

En négligeant le terme $o(x^2, y^2)$ (valable pour les petits déplacements), on obtient l'expression analytique suivante :

$$E(x, y) = Ax^2 + 2Cxy + By^2,$$

avec:

- $A = \frac{\delta I}{\delta x}^2 \otimes w$
- $B = \frac{\delta I}{\delta y}^2 \otimes w$
- $C = \left(\frac{\delta I}{\delta x} \frac{\delta I}{\delta y} \right) \otimes w$

$$E(x, y) = (x, y) \cdot M \cdot (x, y)^t,$$

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix}.$$

M: Matrice d'autocorrélation de l'image I au point (u,v). Elle représente la variation locale de l'image I en (u,v).

Détection des Points d'intérêts

- Harris et Stephen propose l'opérateur suivant pour détecter les coins dans une image :

$$R = Det(M) - kTrace(M)^2$$

$$\text{avec : } Det(M) = AB - C^2 \text{ et } Trace(M) = A + B.$$

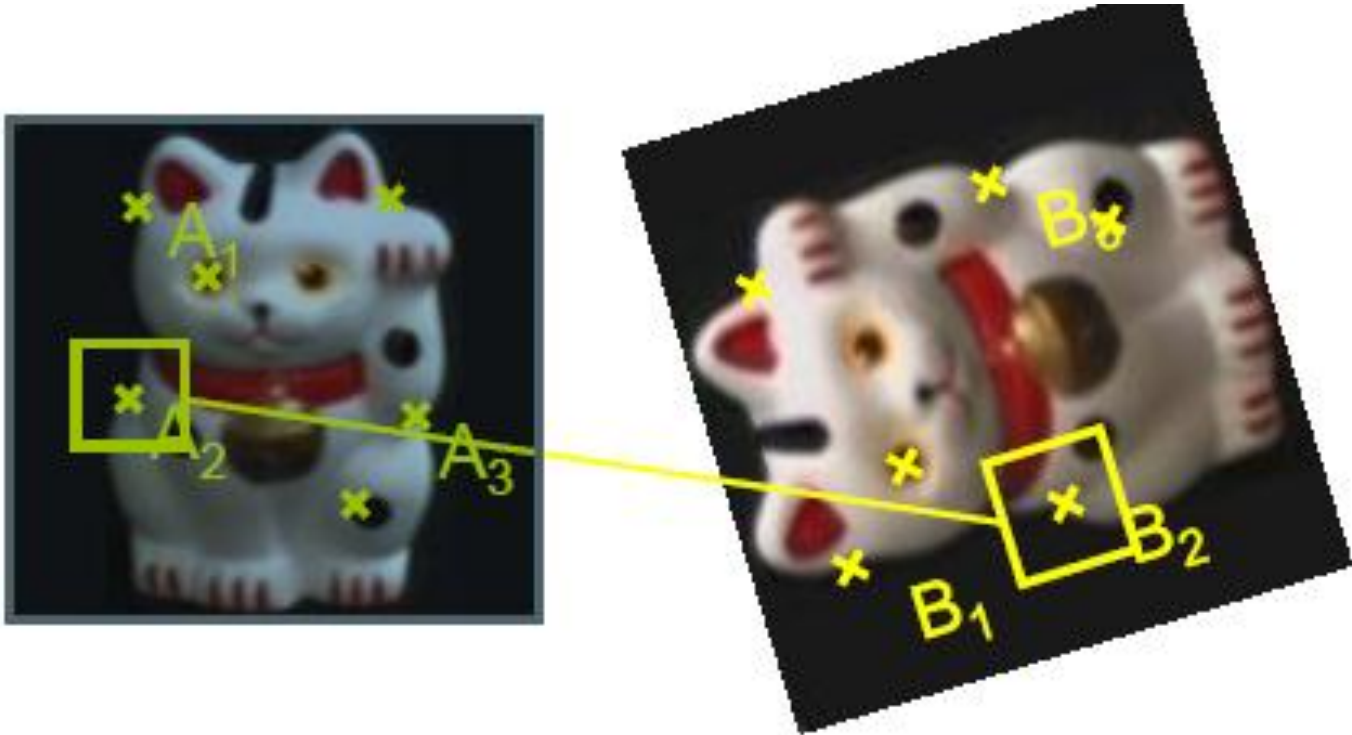
Les valeurs de R sont positives au voisinage d'un coin, négatives au voisinage d'un contour et faibles dans une région d'intensité constante.

Limitations:

Sensible aux changements d'échelle et rotations.



Détection des Points d'intérêts



Descripteur HOG

Descripteur HOG (Histogram of Oriented Gradients)

- Ce descripteur a été présenté par Dalal et Triggs à la conférence CVPR. Il permet de calculer des histogrammes locaux de l'orientation du gradient sur une grille dense en divisant l'image en régions adjacentes de petite taille, appelées cellules, et en calculant pour chaque cellule l'histogramme des orientations des contours pour les pixels à l'intérieur de cette cellule. La concaténation des histogrammes obtenus constitue ce qu'on appelle le descripteur HOG.
- Il est à rappeler que l'amplitude d'un vecteur de gradient dans une région homogène d'une image est égale à zéro et qu'un vecteur de gradient sur un contour a une valeur élevée qui indique l'intensité de ce contour, mais il représente également la direction du contour car il a toujours une orientation perpendiculaire à ce contour.

Descripteur HOG

➤ L'idée de base pour l'utilisation du descripteur HOG est que l'apparence et la forme de l'objet dans une image peuvent être aisément décrites par la distribution des gradients d'intensité et orientation des pixels de cet objet.

Pour une meilleure précision, les histogrammes locaux sont normalisés en contraste sur tout le bloc puis utilisés pour normaliser toutes les cellules du bloc, ce qui aura pour effet une meilleure invariance aux changements d'éclairage et d'ombrage.

Le descripteur HOG présente l'avantage d'être invariant aux transformations géométriques et photométriques mais il est considéré comme coûteux en temps.

Descripteur HOG

Etapes à suivre pour la construction du descripteur HOG

- Calcul du gradient

La méthode la plus utilisée consiste à appliquer un filtre dérivatif du type Sobel . Dallal et Trigs ont testés plusieurs masques complexes.

- Construction de l'histogramme

Construction des histogrammes de l'orientation des gradients répartis sur 0 à 180 degrés ou 0 à 360. Ceci est fait généralement pour des cellules carrées de petite taille de 4x4.

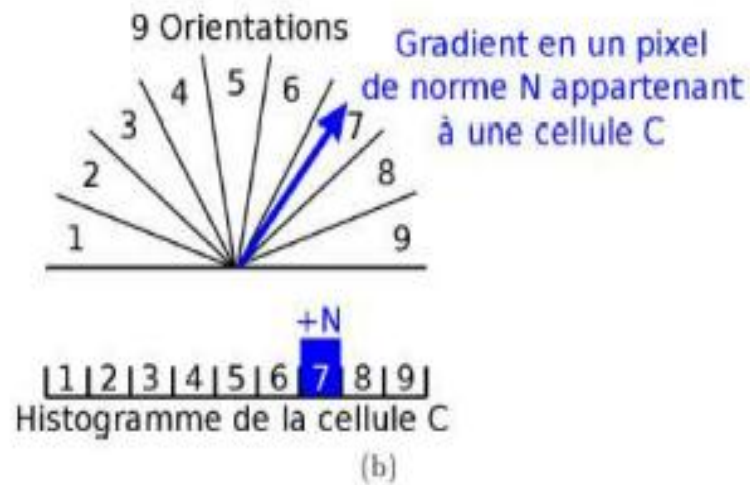
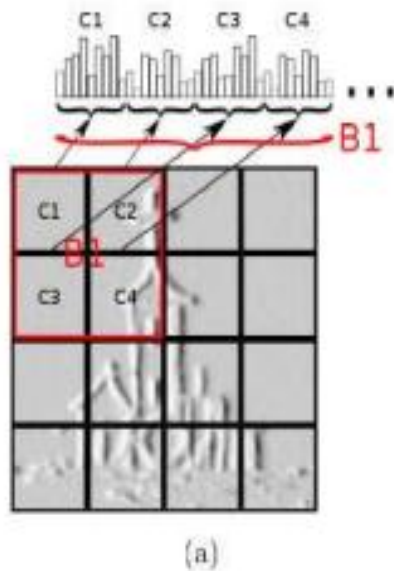
- Concaténation des histogrammes

Cela consiste à juxtaposer les histogrammes obtenues dans l'étape précédente.

Ce descripteur est largement utilisé en vision par ordinateur dans diverses applications car il est insensible à l'orientation et l'échelle.

Descripteur HOG

- Exemple de construction d'un descripteur HOG



Détection des Points d'intérêts

- **SIFT: Scale-Invariant Feature Transform**

En 2004, le chercheur David Lowe a mis au point le descripteur SIFT à l'Université de Colombie Britannique. C'est un descripteur utilisé en vision par ordinateur pour détecter et identifier les éléments similaires entre différentes images numériques (objets, personnes...). Il est très utilisé dans la recherche d'image par le contenu, le tracking vidéo , détection d'objets.....

Le descripteur SIFT permet de caractériser le contenu visuel d'une image de la façon la plus indépendante possible **de l'échelle, de la rotation et du point de vue ainsi que des changement de luminosité.**

Détection des Points d'intérêts

- Calcul du descripteur SIFT

Il s'agit tout d'abord de détecter sur l'image des zones circulaires « intéressantes », ROI (**R**egion **o**f **I**nterest) centrées autour d'un *point d'intérêt* et de rayon déterminé appelé *facteur d'échelle*. Celles-ci sont caractérisées par leur unité visuelle et correspondent en général à des éléments distincts sur l'image.

Détection des Points d'intérêts

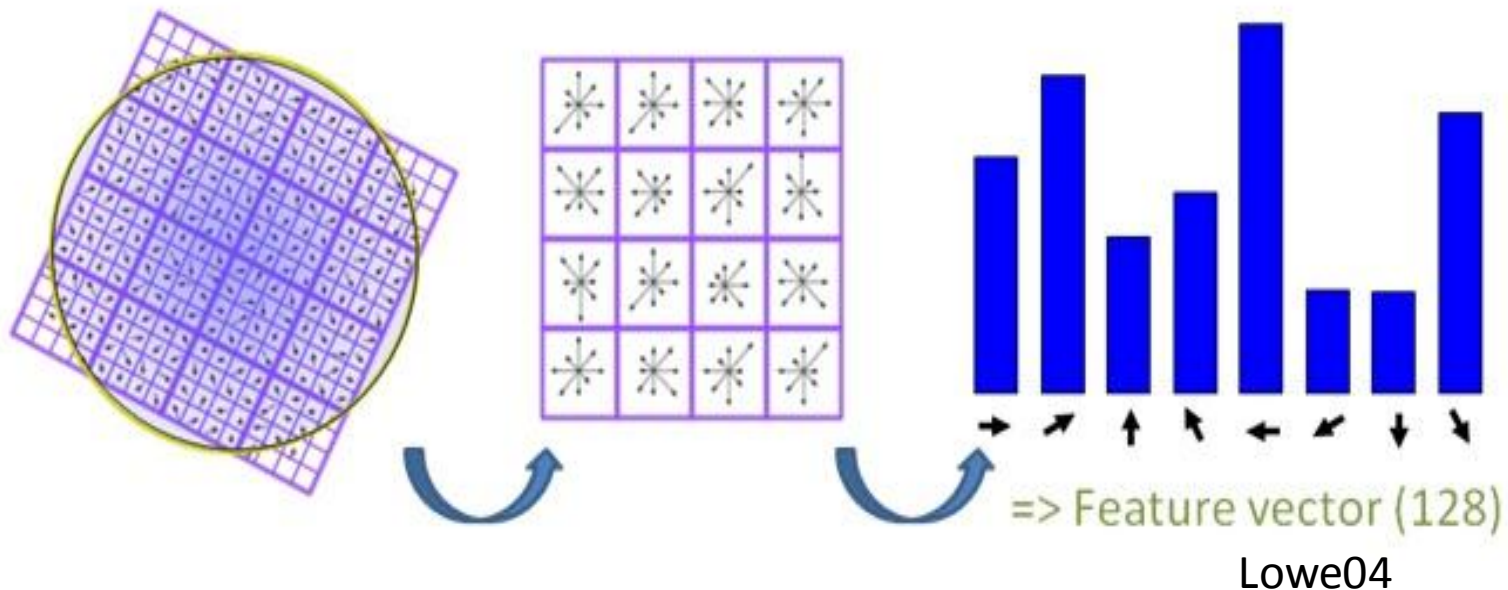
Sur chacune d'elles, on détermine une orientation intrinsèque (c'est-à-dire ne dépendant que du contenu local de l'image autour du point clé), qui sert de base à la construction d'un histogramme des orientations locales des contours (HoG), habilement pondéré, seuillé et normalisé pour plus de stabilité. L'espace autour du point d'intérêt (x,y) est divisé en blocs de 4×4 . Pour chaque bloc, on calcule un histogramme des orientations en 8 directions. Ensuite, les 16 histogrammes à 8 intervalles chacun sont concaténés et normalisés.

C'est cet histogramme qui sous la forme d'un vecteur à 128 ($4 \times 4 \times 8$) valeurs constitue le descripteur SIFT du point-clé, et l'ensemble des descripteurs d'une image établissent ainsi une véritable **signature numérique** du contenu de celle-ci.

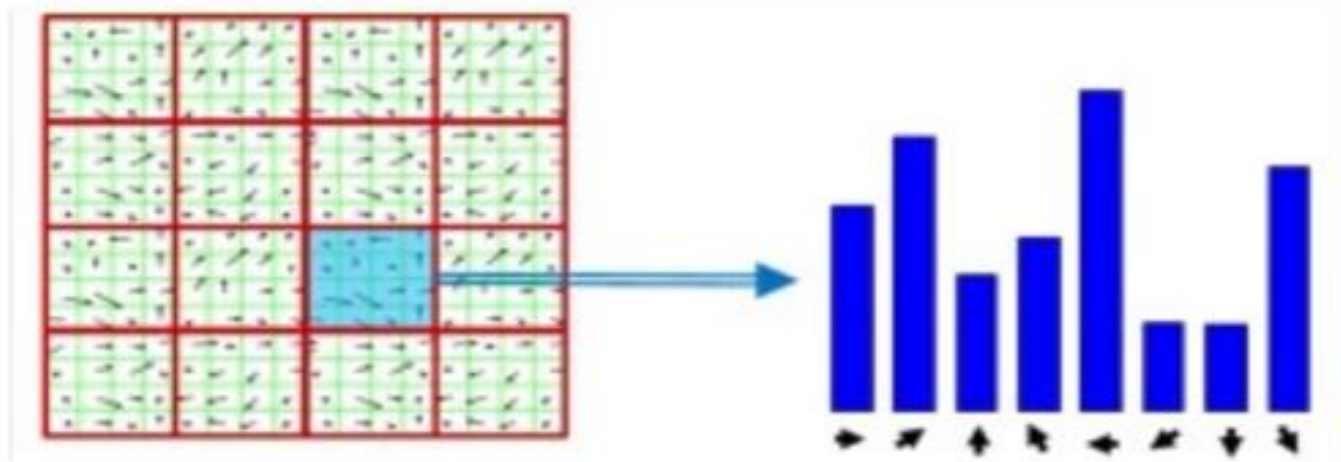
Détection des Points d'intérêts

Ces descripteurs présentent l'avantage d'être invariants à l'orientation et à la résolution de l'image, ainsi qu'au point de vue 3D.

Le descripteur décrit la direction de chaque pixel appartenant au voisinage du point clé en calculant le gradient de chacun et en définissant l'histogramme pour chaque zone de 4×4 .



Détection des Points d'intérêts



- Concaténer les 16 histogrammes pour obtenir un descripteur de 16x8



Détection des Points d'intérêts

Descripteur SIF= histogrammes sous la forme d'un vecteur normalisé (de dimension 128)

Un peu plus récemment (2006 et 2008), **SURF**: **S**peeded **U**p **R**obust **F**eatures: une variante de SIFT 6 fois plus rapide a été proposé en 2006, à la conférence ECCV'06.

Détection des Points d'intérêts

En Résumé

Pour chaque pixel de l'image, on calcule le vecteur gradient. Le gradient nous donne des informations sur le contour local.

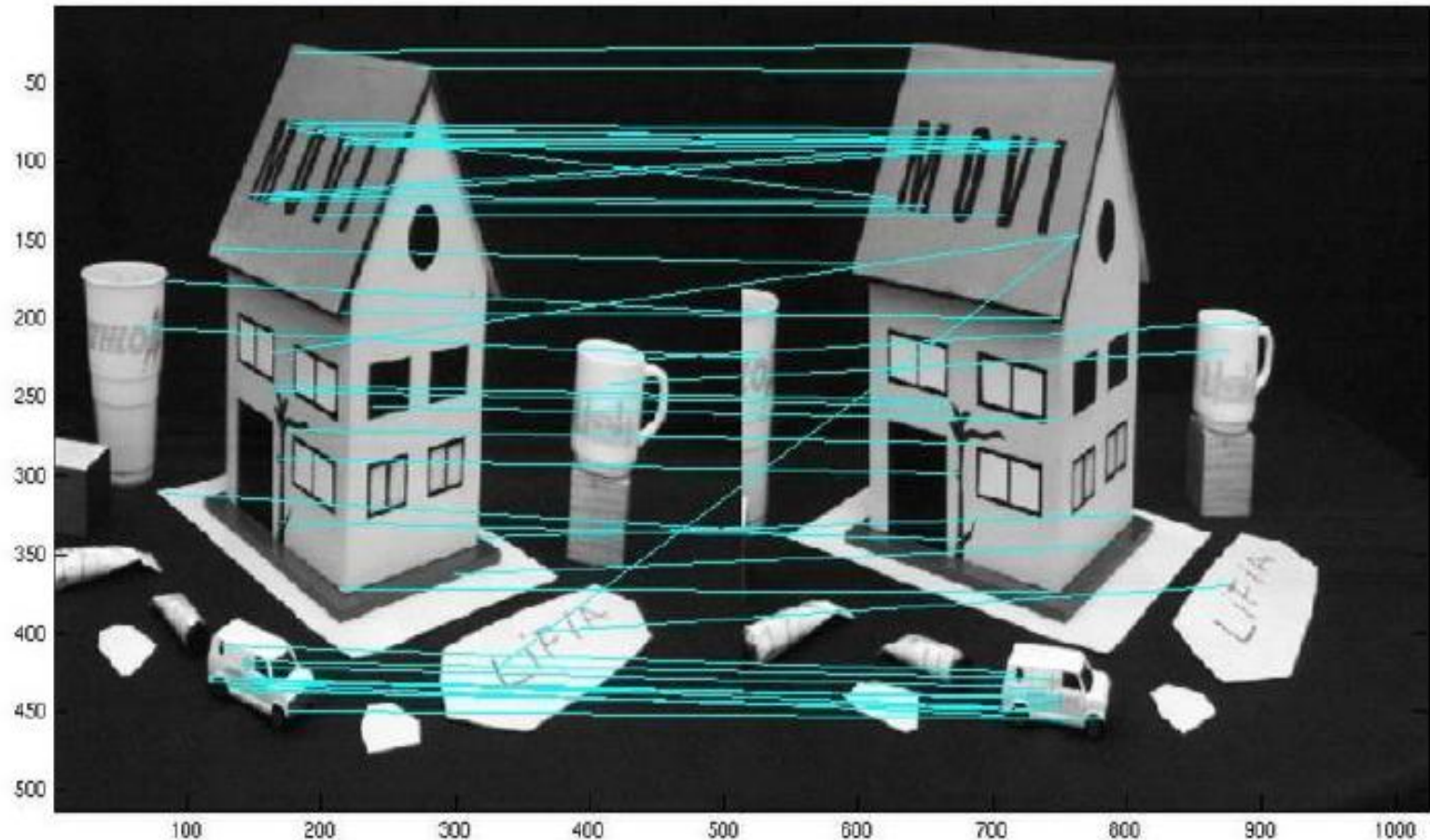
Les keypoint(points d'intérêt) dans SIFT sont des "coins", c'est à dire des endroits où le contour change brutalement d'orientation. Ce sont des points particulièrement intéressants car ils sont très spécifiques, il ne sont pas nombreux.

Un bon moyen de décrire ce qui se passe autour d'un keypoint, c'est de décrire les contours autour du keypoint. Donc de regarder les vecteurs gradient dans le voisinage : est-ce qu'ils vont tous dans la même direction, ou dans 2 directions principales, ou dans tous les sens,...

Pour cela on construit un histogramme qui représente la répartition des vecteurs gradient dans les 8 directions. Lowe a choisi de faire 16 histogrammes, chacun représentant un voisinage de 4x4 pixels.

En mettant bout à bout toutes ces informations, on obtient une "signature" composée de $16 \text{ histogrammes} \times 8 \text{ directions} = 128 \text{ valeurs}$.

Appariement des points SIFT



Résultat d'appariement par SIFT entre l'image (2) à gauche, 510 points détectés, et l'image (1) à droite, 589 points détectés. 51 points ont été appariés, ce qui correspond à une distance euclidienne entre les descripteurs en deçà d'un certain seuil.

[Lowe 04]

Détection des Points d'intérêts

- Avantages des points d'intérêts :
 - Sources d'informations plus fiable que les contours car plus de contraintes sur la fonction d'intensité.
 - Robuste aux occultations (soit occulté complètement, soit visible).
 - Présents dans une grande majorité d'images.

Sac de mots BOW

- Dans ce modèle, un texte (tel qu'une phrase ou un document) est représenté comme le sac (multiset) de ses mots, sans tenir compte de la grammaire ni même de l'ordre des mots, mais en conservant la multiplicité .
- En 2003, le modèle du sac de mots a également été utilisé pour la vision par ordinateur .
- Le modèle de sac de mots est couramment utilisé dans les méthodes de classification.

Sac de mots BOW

□ Descripteur BOW: Bag of Words(sac de mots visuels)

- Le descripteur BOW permet de décrire une image comme un ensemble de mots visuels appelés également vocabulaire visuel. C'est une technique permettant de décrire de manière compacte des images.
- Le vocabulaire est obtenu en regroupant des entités extraites de l'image correspondant à des mots visuels. D'une autre manière, une image est représentée par un histogramme où chaque bin correspond à un mot visuel et le poids associé représente l'importance du mot dans l'image.

Sac de mots BOW

Construction du descripteur BOW:

- Extraire les caractéristiques visuelles,
- Construire un vocabulaire visuel

- Extraire les caractéristiques visuelles:

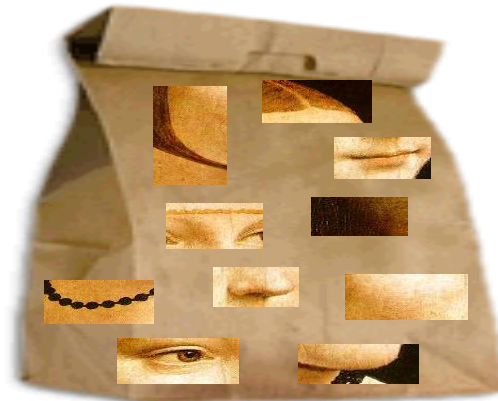
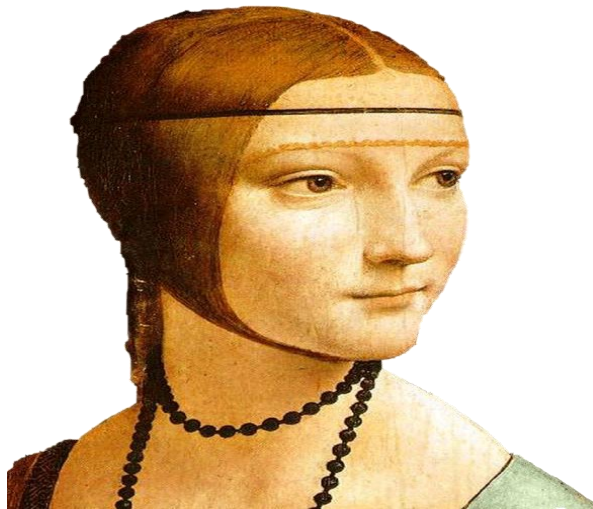
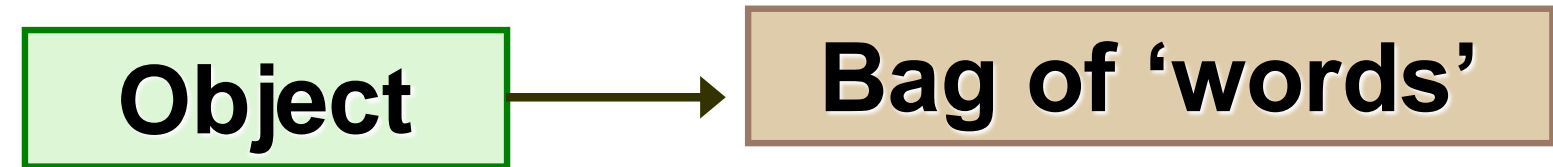
Dans cette étape, les caractéristiques locales des images sont extraites en utilisant plusieurs approches comme extraction des points SIFT.

Sac de mots BOW

- Construire le vocabulaire visuel:

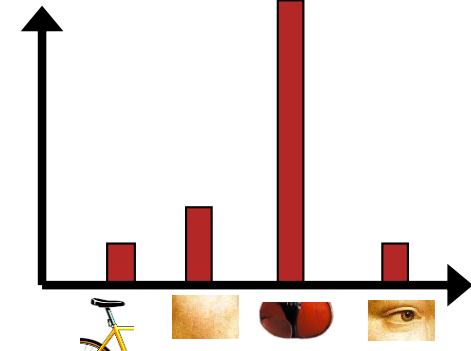
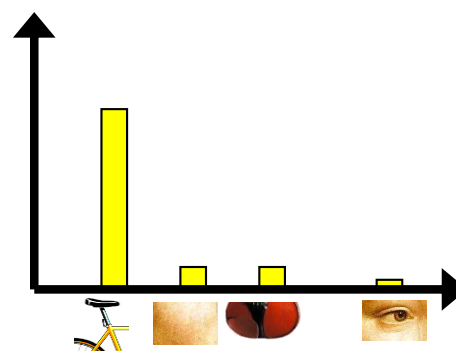
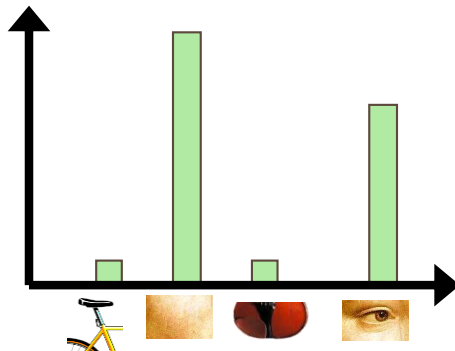
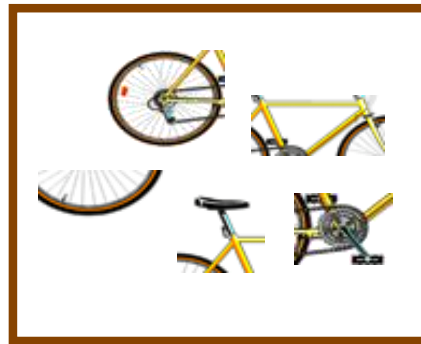
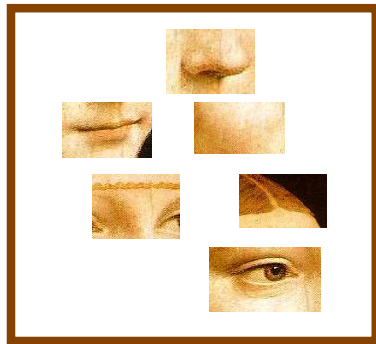
Le vocabulaire peut être généré par exemple en retrouvant les clusters autour des points SIFT en utilisant par exemple l'algorithme Kmeans. La taille du vocabulaire va être le nombre de clusters.

Exemple



Les étapes de Bag-of-words

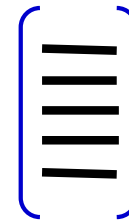
1. Extraire les caractéristiques de l'objet,
2. Apprentissage du vocabulaire visuel,
3. Quantification des caractéristiques en utilisant le vocabulaire visuel,
4. Représenter les images par les fréquences des mots visuels “visual words”



Extraire les caractéristiques de l'objet

Extraction des mots visuels:

Correspondre à chaque mot visuel un **descripteur**



Descripteur

Un descripteur peut être décrit par plusieurs attributs exemple:

- La couleur,
- Histogramme d'orientation :HOG
- La forme (la silhouette),
- Les points caractéristiques: Harris, SIFT, SURF,
- Les moments mathématiques tels que Tchebychev, Zernike.....

Problème: Ne distingue pas le foreground (avant plan) du background (arrière plan).

Résumé

- **Problème :**

- Image initiale est généralement dégradée:
(bruit, faible contraste, illumination variable, ...)

Nécessite

- **Prétraitements** pour améliorer l'image initiale
- **Post-traitements** de l'image Résultat

Résumé

- **Prétraitements**

- ✓ **Amélioration du contraste**

Outils : utilisation de l'**histogramme**

- ✓ **Réduction du bruit**

Outils : **filtres de voisinage**

- **Post-traitements**

Outils : **filtres morphologiques**

- ✓ **Élimination des**

- Trous dans les objets
- Artefacts dans le fond

- ✓ **Fusion d'objets scindés**

- ✓ **Division d'objets fusionnés**

Résumé

- **Filtre Passe-bas** : diminue le bruit mais atténue les détails de l'image .
- **Filtre Passe-haut** : accentue les contours et les détails de l'image mais amplifie le bruit.
- **Filtre Passe-bande** : élimine certaines fréquences indésirables présentes dans l'image.

Résumé

- L'idée derrière le filtrage Linéaire (Gaussien ou le filtre moyennneur) est que des pixels voisins dans l'image ont des illuminances voisines.
- On fait l'hypothèse que l'illuminance varie lentement et qu'un moyennage spatial permet de réduire le bruit.
- Si l'hypothèse de variation de l'illuminance faible n'est pas respectée (comme sur les arêtes des images correspondant aux changements rapides de l'illuminance), le moyennage adoucit ces discontinuités, ce qui n'est pas le but recherché.
- En vision artificielle, l'utilisation de la transformée de Fourier est peu répandue car il est plus naturel de travailler directement dans l'image.