

AIV2 – Description globale des images statiques

Pierre Tirilly

Master Informatique, parcours RVA – Université de Lille

Dans des épisodes précédents...

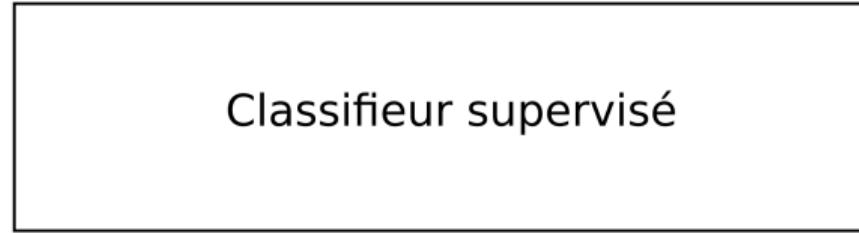
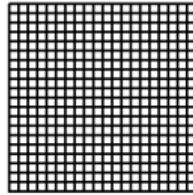
Traitement / analyse d'images (IMG / AIV1) :

- ▶ Histogramme de niveaux de gris
- ▶ Convolution
- ▶ Caractérisation des textures
- ▶ Points d'intérêt / caractéristiques locales (SIFT, etc.)
- ▶ Segmentation d'images
- ▶ Clustering

Machine learning (AIV2 / DLV) :

- ▶ Classification supervisée
- ▶ Réseaux de neurones / deep learning

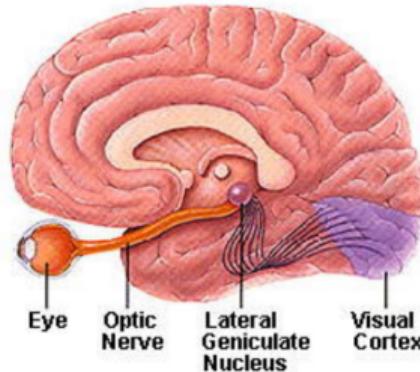
Dans le dernier épisode d'AlV2...



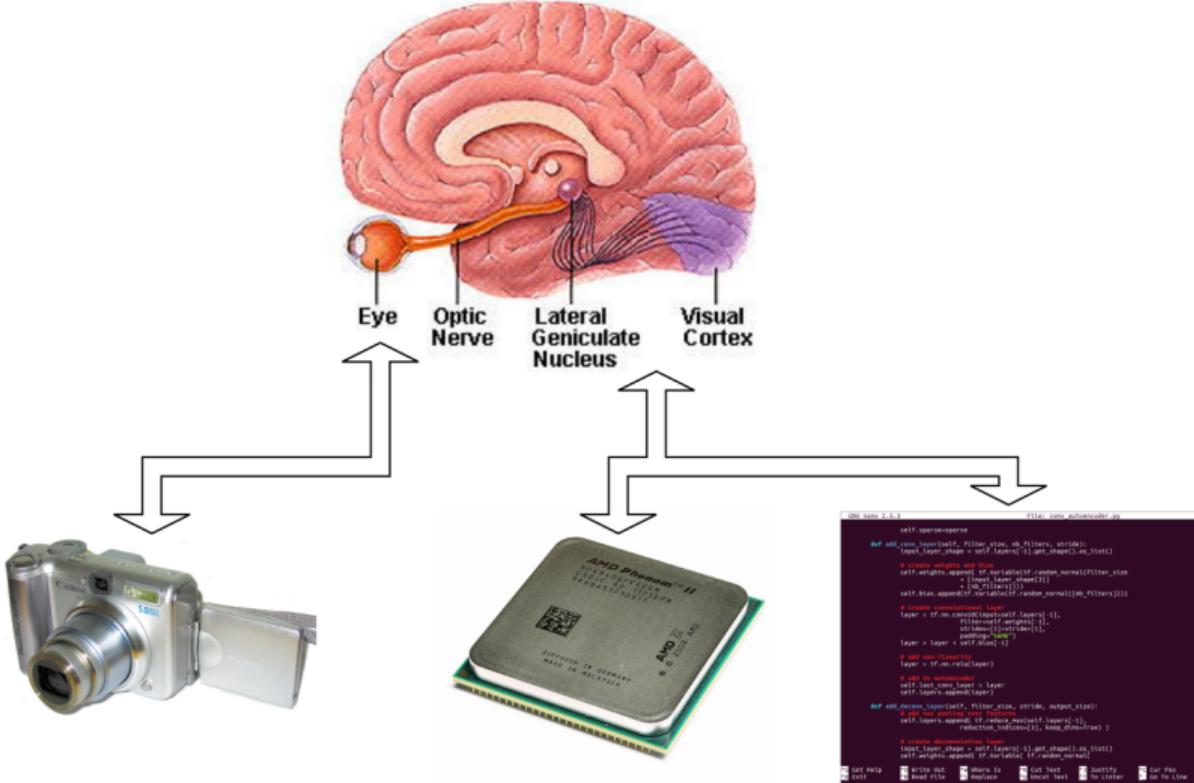
3

Vision par ordinateur

Vision par ordinateur : Qu'est-ce que c'est ?



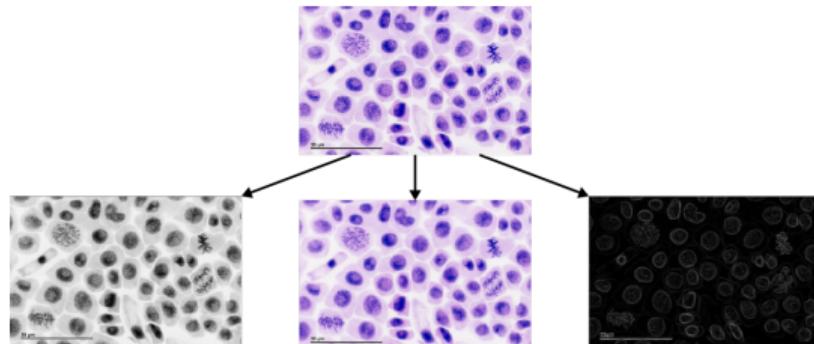
Vision par ordinateur : Qu'est-ce que c'est ?



Vision par ordinateur vs. Traitement d'images

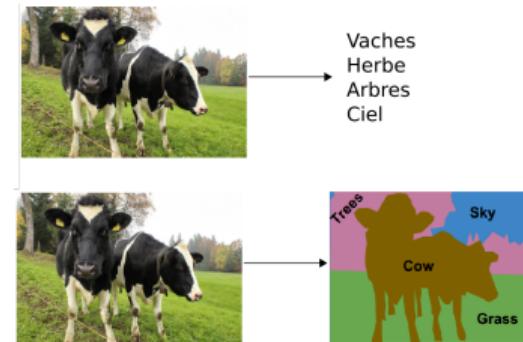
Traitement d'images : *transformation* des images à l'aide d'algorithmes

- ▶ Entrée : Image
- ▶ Sortie : Image



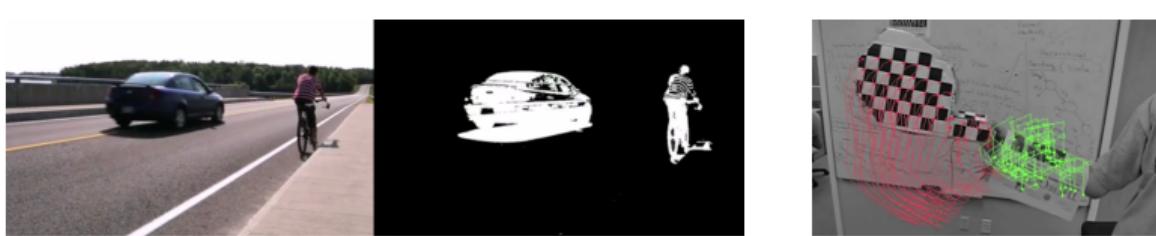
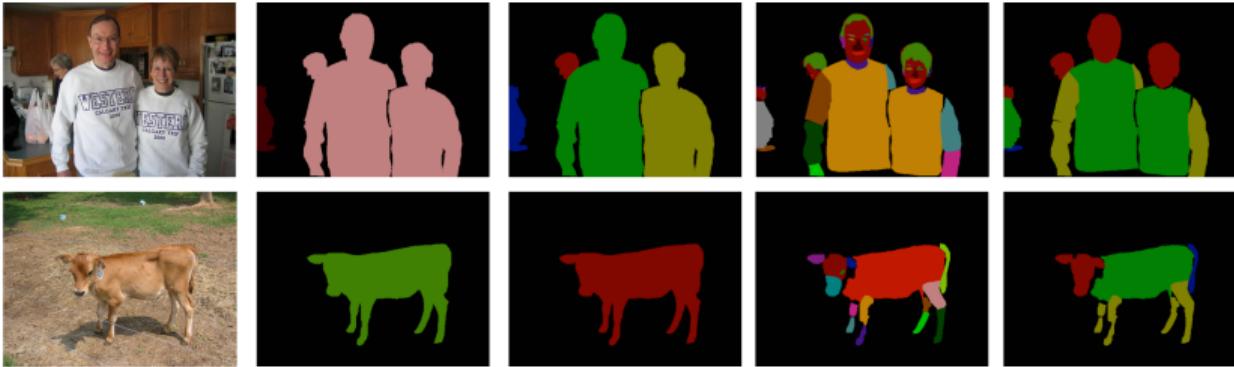
Vision par ordinateur : *compréhension* des images à l'aide d'algorithmes

- ▶ Entrée : Image
- ▶ Sortie : Connaissance

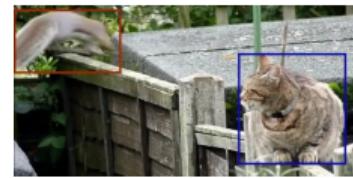
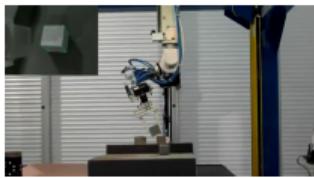
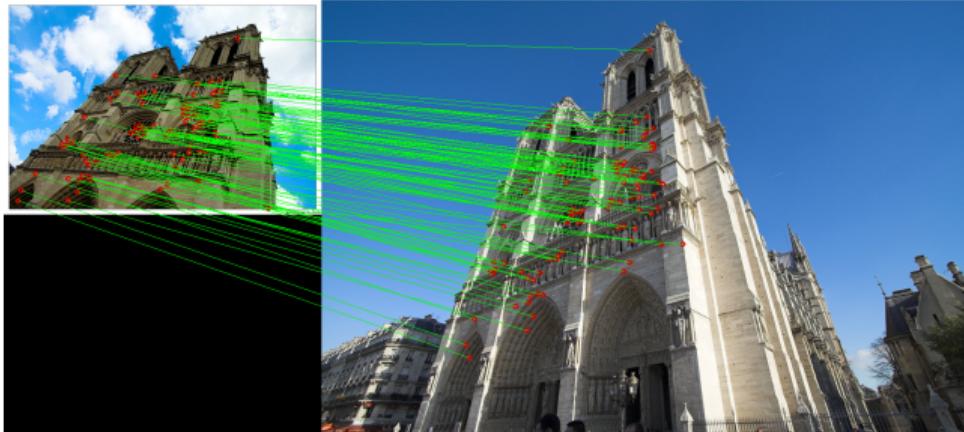


Source : tariq-hasan.github.io

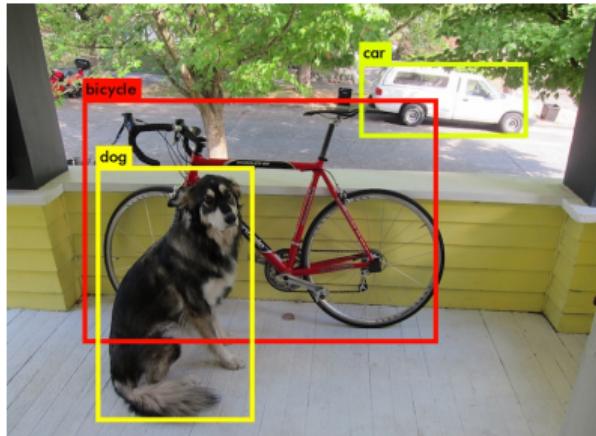
(Ré)organisation



Mise en correspondance

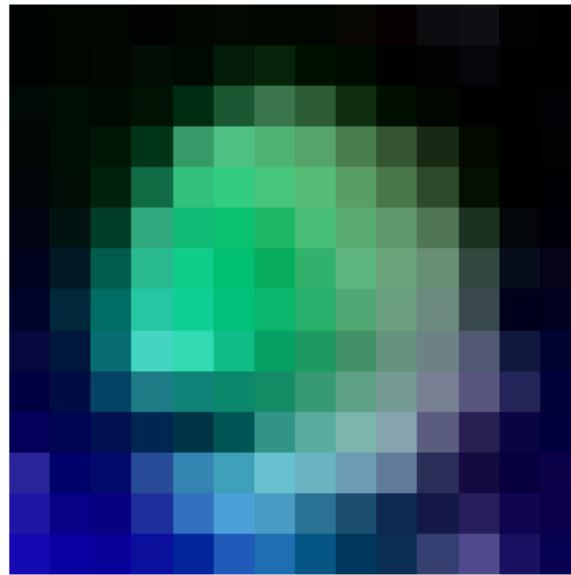


Reconnaissance



Est-ce difficile ?

Puissance de la vision humaine



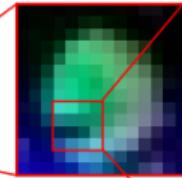
Puissance de la vision humaine



Puissance de la vision humaine



Ce que voit la machine

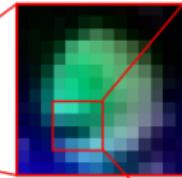


67	49	16	5
30	15	11	19
0	0	0	50
40	52	62	104

213	218	189	161
124	131	135	139
39	53	86	148
76	133	160	193

194	177	135	100
134	120	107	101
82	70	84	135
152	178	185	207

Ce que voit la machine



67	49	16	5
30	15	11	19
0	0	0	50
40	52	62	104

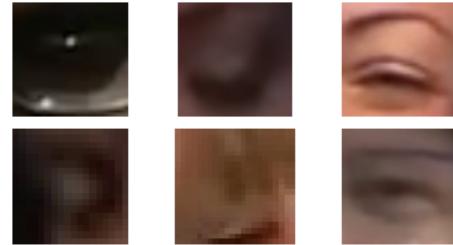
213	218	189	161
124	131	135	139
39	53	86	148
76	133	160	193

194	177	135	100
134	120	107	101
82	70	84	135
152	178	185	207

Variabilité visuelle



Variabilité visuelle



Ambiguité visuelle



Ambiguité visuelle



Ambiguité visuelle



Reconnaissance d'objets

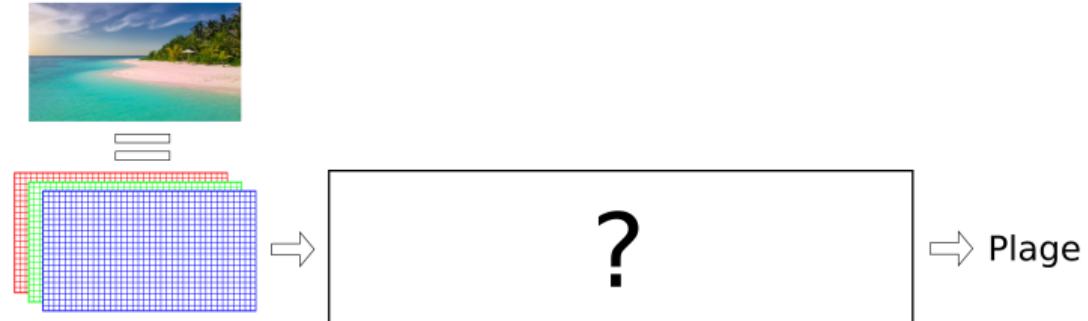
Formaliser le problème

Entrée : Image

- ▶ En niveaux de gris : Matrice $I \in M_{hw}$, $I[i, j] \in [0, 255]$ ou $I[i, j] \in [0, 1]$
- ▶ En couleur : Matrices (tenseur) $I \in M_{hw}^3$, $I^c[i, j] \in [0, 255]$ ou $I^c[i, j] \in [0, 1]$

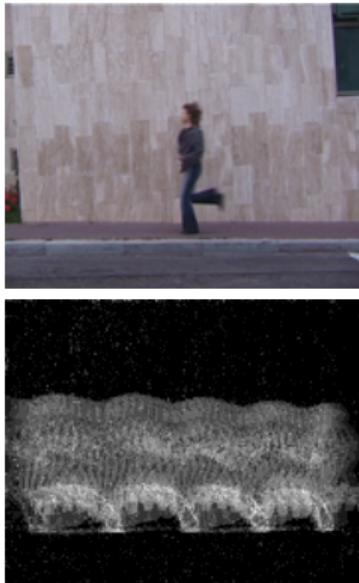
Sortie : Classe = valeur discrète (ou *nominal*) c dans un ensemble fini $\mathcal{C} = \{c_1, c_2, \dots, c_{|\mathcal{C}|}\}$

Objectif : Établir une fonction $g : \mathcal{I} \rightarrow \mathcal{C}$ associant sa classe à une image

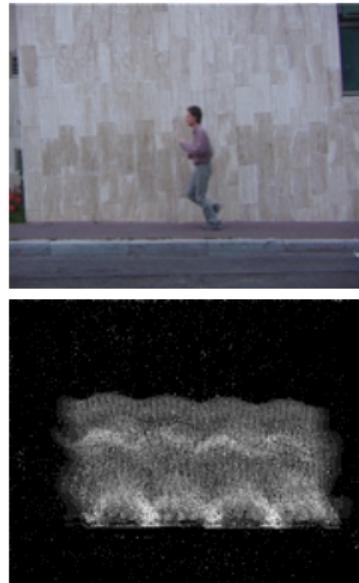


Difficultés spécifiques au problème de reconnaissance

Variance intra-classe



Courir



Courir



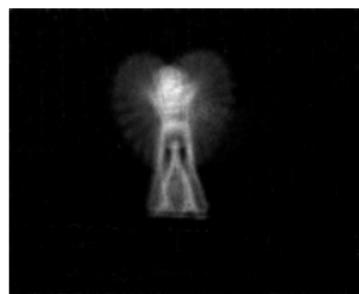
Courir

Difficultés spécifiques au problème de reconnaissance

Variance inter-classe



Jump



Jack

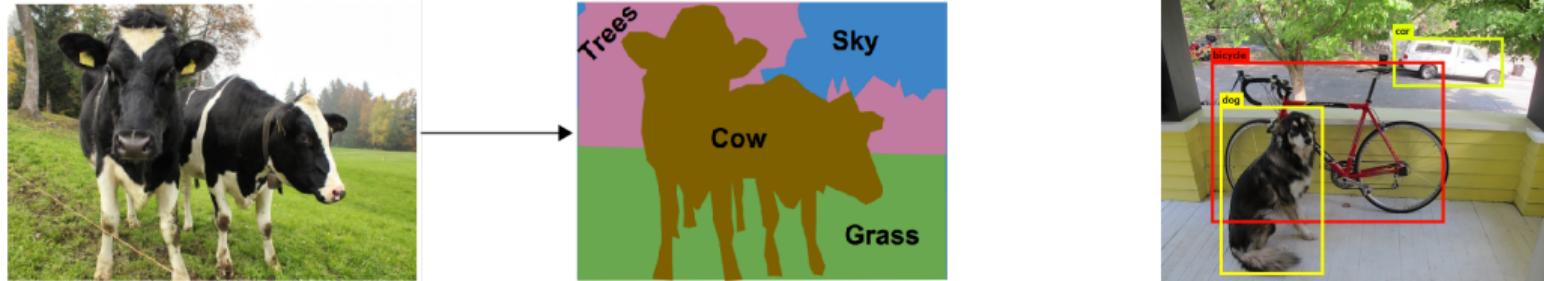


Wave

Vision par ordinateur et machine learning

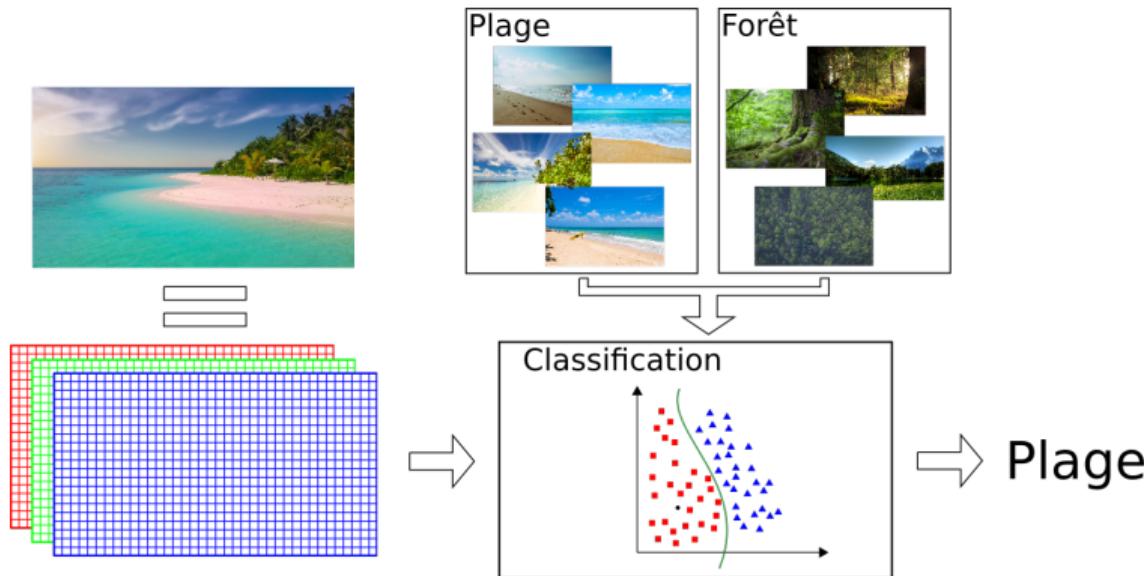
Constat : Les problèmes de vision par ordinateur peuvent se formuler comme des problèmes de machine learning

- ▶ Reconnaissance, segmentation sémantique → Classification supervisée
- ▶ Détection, suivi → Régression (supervisée)
- ▶ Segmentation → Clustering
- ▶ ...



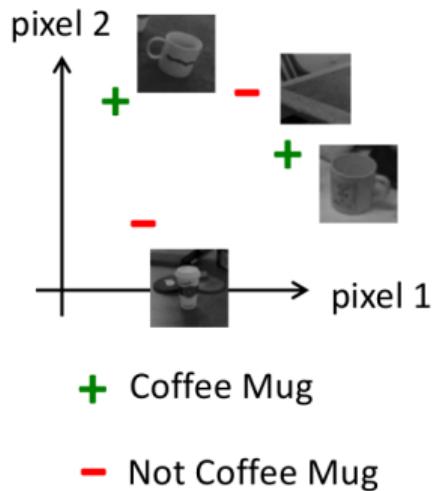
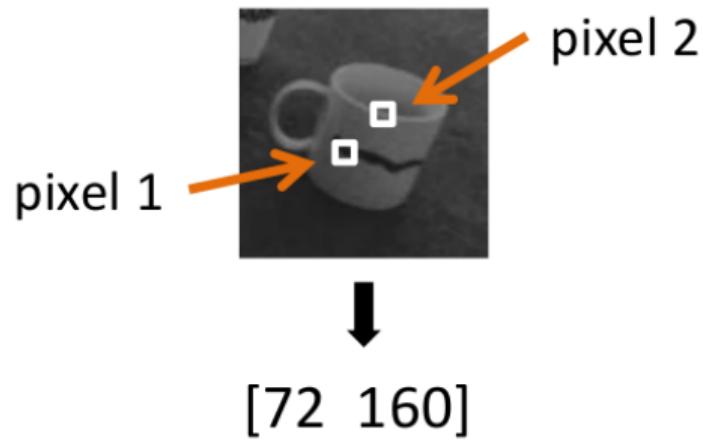
Limites du machine learning

Mais : Les problèmes de vision par ordinateur sont trop complexes pour une approche “naïve” de machine learning



Limites du machine learning

Mais : Les problèmes de vision par ordinateur sont trop complexes pour une approche “naïve” de machine learning



Source : [A. Coates]

Programme

Programme des six prochaines séances

Problématique : Comment représenter l'information visuelle (images/vidéos) pour des tâches de vision par ordinateur ?

Partie 1 : Images statiques

- ▶ Séance 1 : Représentation globale des images
- ▶ Séance 2 : Représentation locale des images
- ▶ Séance 3 : Apprentissage de représentations
- ▶ Application : Reconnaissance d'objets

Partie 2 : Vidéos

- ▶ Séance 4 : Extraction du mouvement
- ▶ Séance 5 : Représentation du mouvement
- ▶ Application : Reconnaissance d'actions

Évaluation (séance 6) : nano-hackaton

Descripteurs visuels

Descripteur visuel

Descripteur visuel : Représentation d'une image (ou région d'image) qui soit

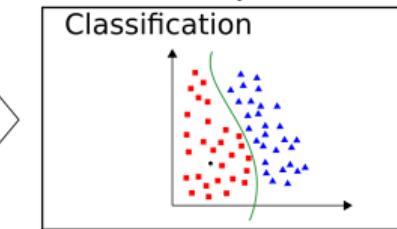
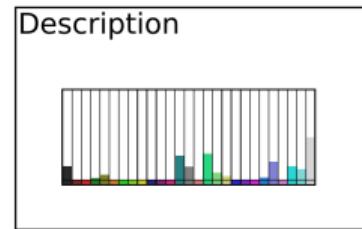
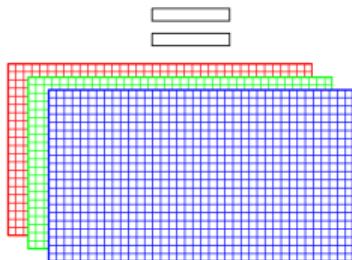
- ▶ Compacte
- ▶ Représenteative de son contenu visuel
- ▶ (En général) Sous forme d'un vecteur dans \mathbb{R}^d

Objectif :

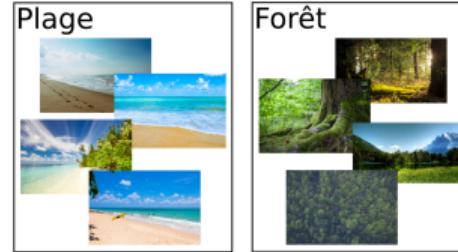
- ▶ Images / objets similaires \leftrightarrow descripteurs proches (au sens mathématique)
- ▶ Images / objects différents \leftrightarrow descripteurs éloignés
- ▶ Classe d'objets séparable (idéalement, linéairement) dans l'espace de description

Synonymes : caractéristique (visuelle), $\{\text{visual}, \text{image}\}$ feature, $\{\text{visual}, \text{image}\}$ descriptor

Reconnaissance avec descripteurs

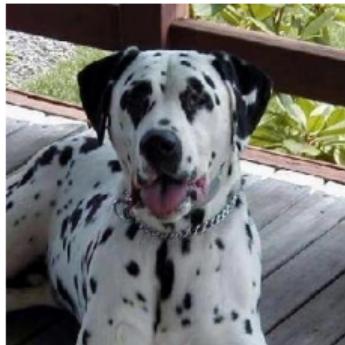


Plage



Quelle information visuelle représenter ?

- ▶ Couleur : distribution des couleurs (intensité, teinte, saturation)
- ▶ Texture : motifs répétés dans les images
- ▶ Forme : contours ou éléments caractéristiques des objets

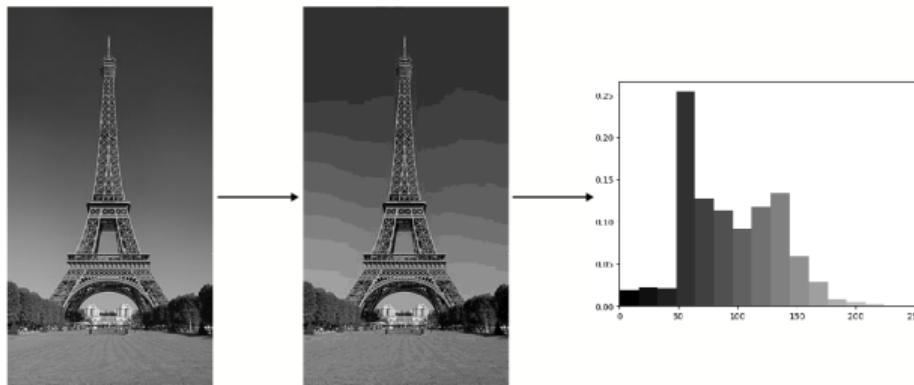


Description des couleurs

Histogramme de niveaux de gris

Histogramme : Représente les proportions de niveaux de gris dans une image

- ▶ [Optionnel] Réduction du nombre de niveaux de gris (*quantification*) : $I_q[i,j] \leftarrow \lfloor \frac{I[i,j]}{w} \rfloor * w$
avec $w = \frac{256}{n_b}$ le pas de quantification, n_b le nombre de *bins* (niveaux de gris)
- ▶ Comptage du nombre de pixels par niveau de gris
- ▶ Normalisation par la taille de l'image → proportions de pixels de chaque niveau de gris

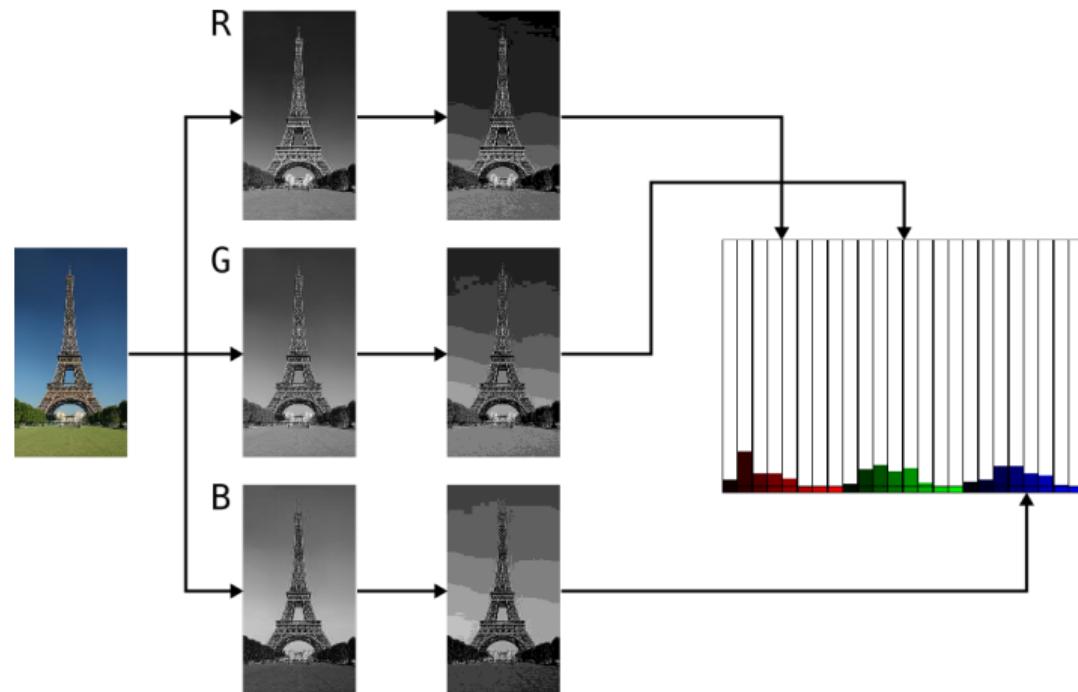


Couleurs : histogramme marginal

Histogramme marginal :

histogramme des canaux de couleurs

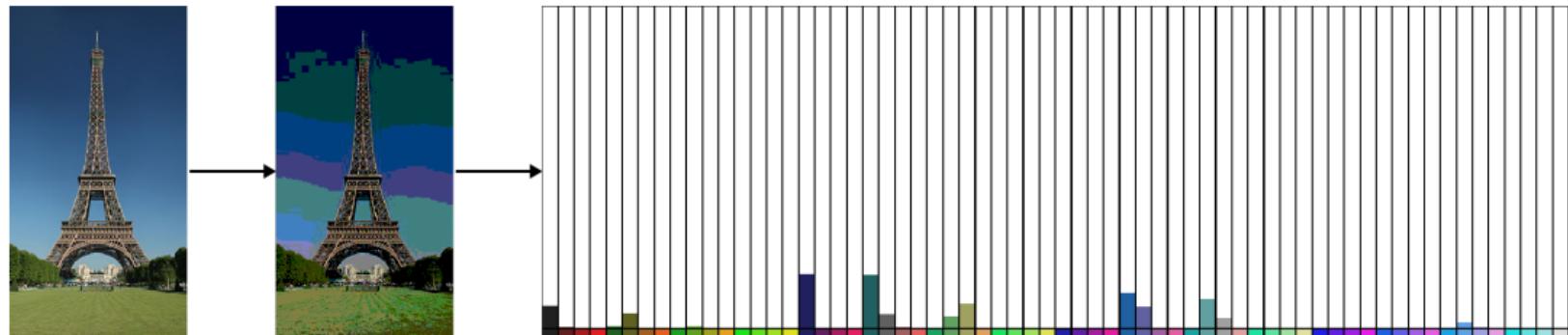
- ▶ Calcul d'un histogramme par canal de couleurs
- ▶ *Concaténation* des histogrammes en un unique histogramme
- ▶ Taille (en bins) de l'histogramme : $3 * n_b$



Couleurs : histogramme joint

Histogramme joint : histogramme de (vraies) couleurs

- ▶ Quantification jointe des couleurs : $I_q[i, j] \leftarrow (\lfloor \frac{I[i, j, 0]}{w} \rfloor * w, \lfloor \frac{I[i, j, 1]}{w} \rfloor * w, \lfloor \frac{I[i, j, 2]}{w} \rfloor * w)$
- ▶ Taille (en bins) de l'histogramme : n_b^3



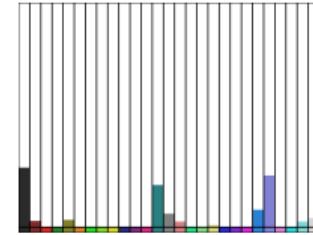
Comparaison d'histogrammes

Mathématiquement : Histogramme = Vecteur

Distances entre vecteurs (exemples) :

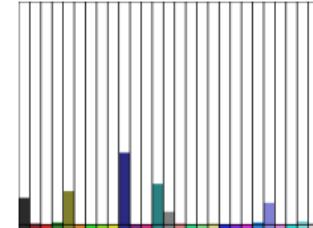
- ▶ Distance euclidienne

$$d_2(h_1, h_2) = \sqrt{\sum_i (h_1[i] - h_2[i])^2}$$

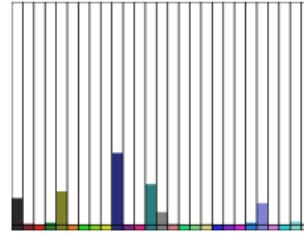
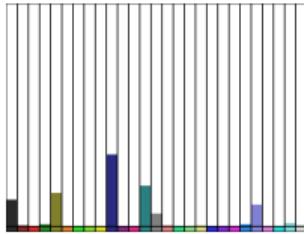
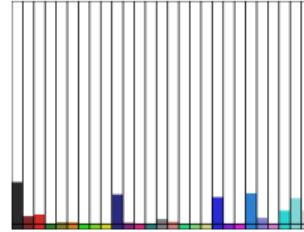
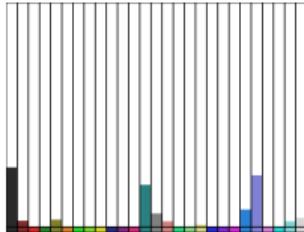


- ▶ Distance de Manhattan

$$d_1(h_1, h_2) = \sum_i |h_1[i] - h_2[i]|$$



Comparaison d'histogrammes

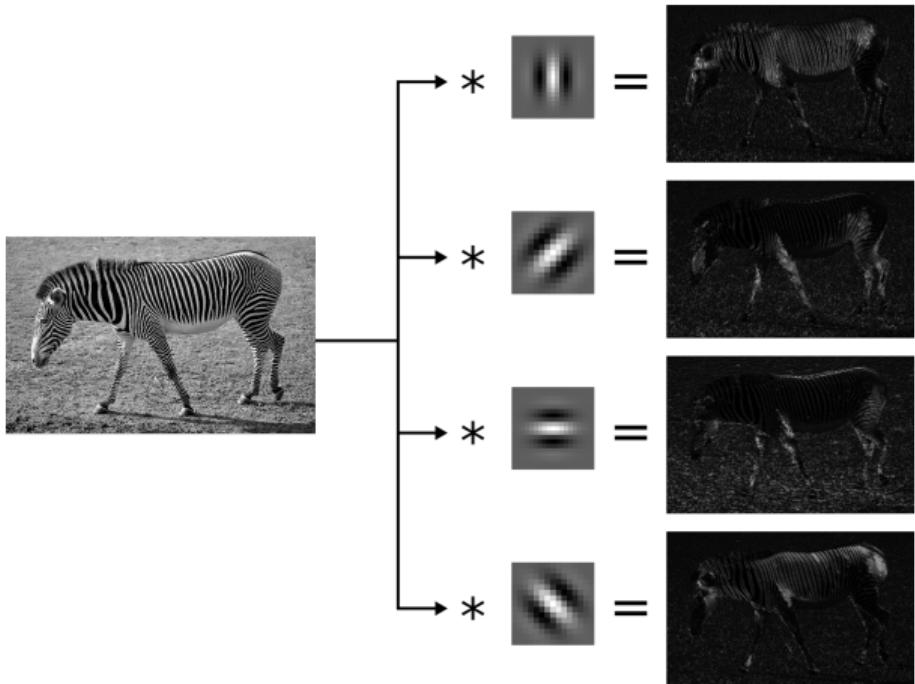


Description des textures

Détection des textures : banques de filtres

Détection de textures : Convolution par des filtres représentant des motifs précis

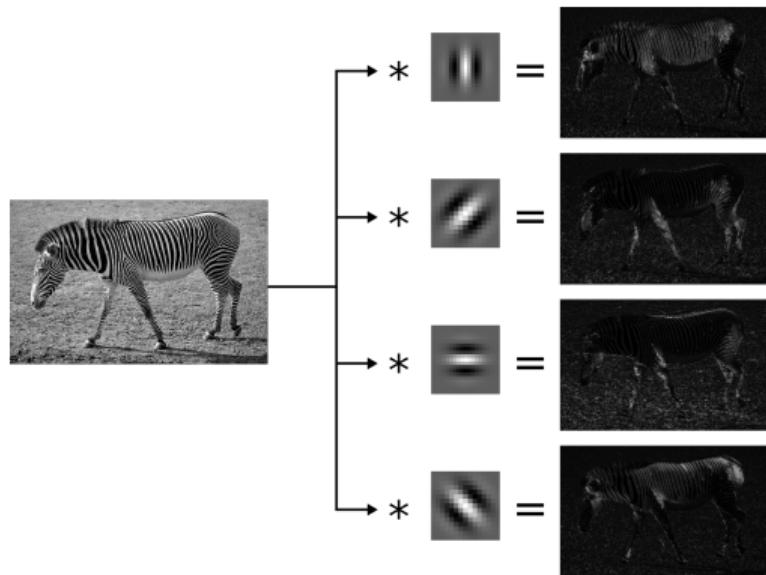
- ▶ Motifs variés en échelle, orientation, fréquence
- ▶ La réponse (image résultante) à un filtre est proportionnelle à la présence du motif à chaque position
- ▶ Exemples : filtres de Gabor, *steerable filters*, etc.



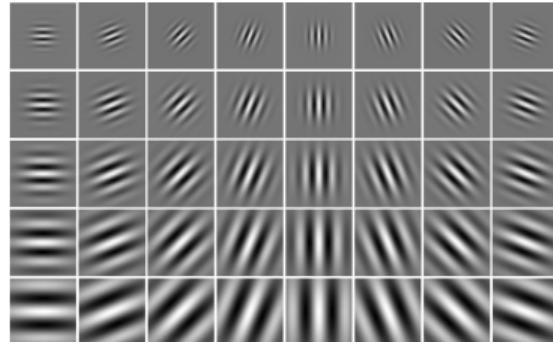
Description des textures : banques de filtres

Description des textures : Calcul de statistiques sur les réponses aux filtres

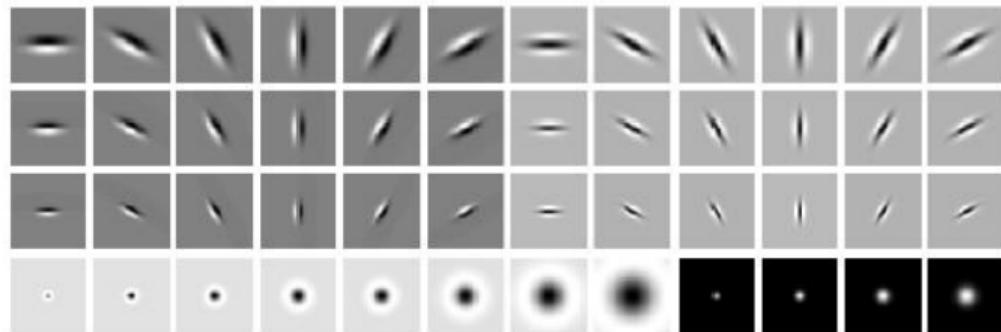
- ▶ Exemple de statistiques : moyenne et écart-type
 - ▶ Calcul des statistiques pour chaque filtre de la banque
- Descripteur = vecteur représentant les statistiques sur l'ensemble des filtres



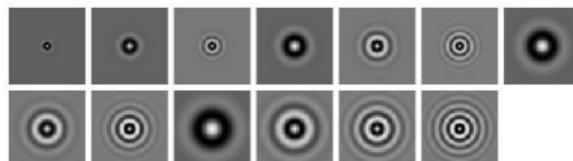
Exemples de banques de filtres



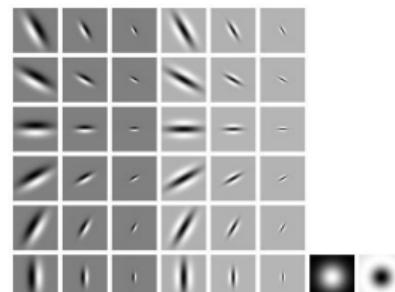
Gabor filters



Leung-Malick filters



Schmid filters

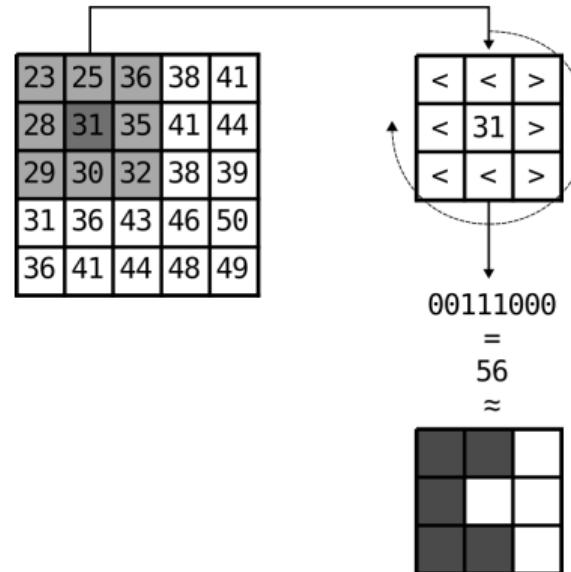


Maximum response filters

Description des textures : Local binary patterns (LBP)

Local binary pattern : signature binaire décrivant le voisinage d'un pixel

- ▶ Comparer le pixel à chacun de ses voisins
- ▶ Codage binaire en fonction du signe de la différence (+ : 1, - : 0)
- ▶ Signature du pixel : concaténation des bits (code sur 8 bits)



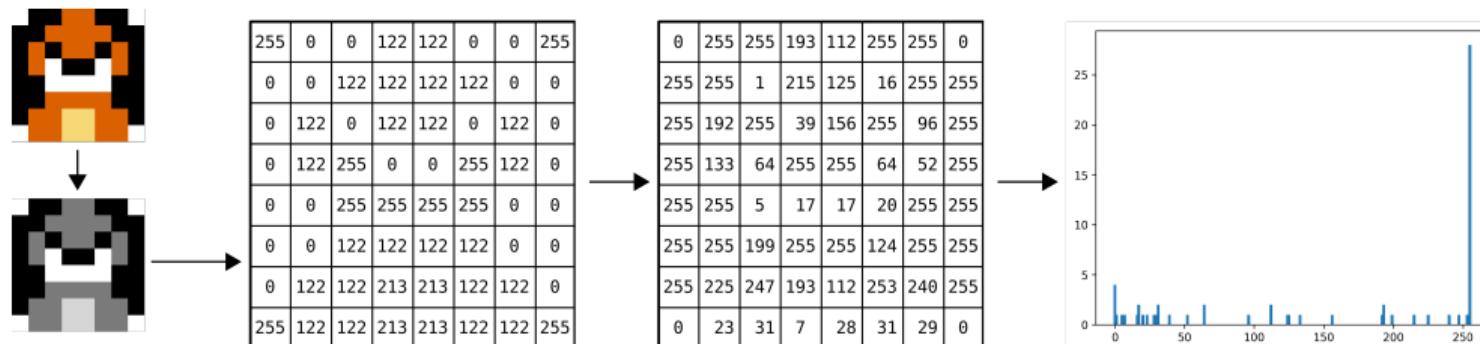
Propriétés :

- ▶ Invariant à l'illumination
- ▶ 256 patterns possibles (pour cette version)

Description des textures : Local binary patterns (LBP)

Descripteur basé LBP : histogramme des LBP de l'image

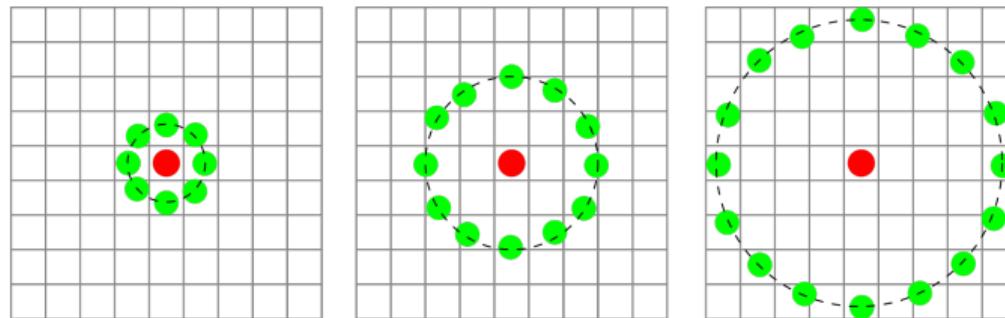
1. Convertir l'image en niveaux de gris
2. Calculer le LBP en chaque pixel
3. Compter le nombre d'occurrences de chaque LBP dans l'image



Local binary patterns (LBP) : Généralisation

Paramètres des LBP :

- ▶ Nombre de voisins
- ▶ Rayon : distance du pixel central à ses voisins



[Source : Wikimedia]

Propriété : Le nombre de voisins n détermine le nombre de patterns / la taille de l'histogramme : 2^n

Description “pseudo-locale” des images

Description globale et ses limites

Descripteur global : Représentation d'une image calculée sur la totalité de ses pixels

Problème : Perte de la structure géométrique de l'image

- ▶ Les objets sont "mélangés" dans le descripteur

Idée : Découper l'image en régions

- ▶ Calcul d'un descripteur par région
- ▶ Constitution d'un descripteur global à partir de ces descripteurs "locaux"

Questions :

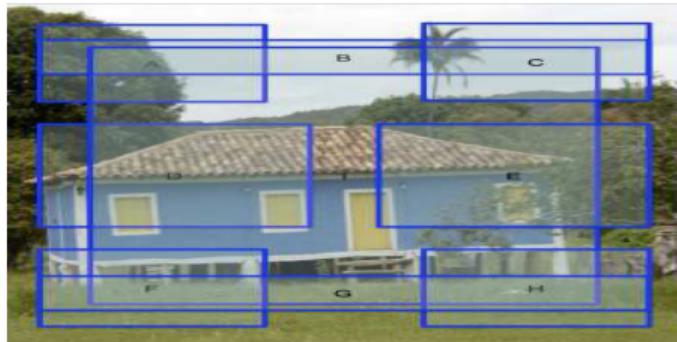
- ▶ Comment choisir les régions ?
- ▶ Comment construire le descripteur final ?



Description locale géométrique : découpage

Approches courantes :

A prioris géométriques



[Tollari et al., 2006]

Grille



Bandes



Variantes :

- ▶ Grille / bandes avec recouvrements
- ▶ Grille / bandes multi-échelles

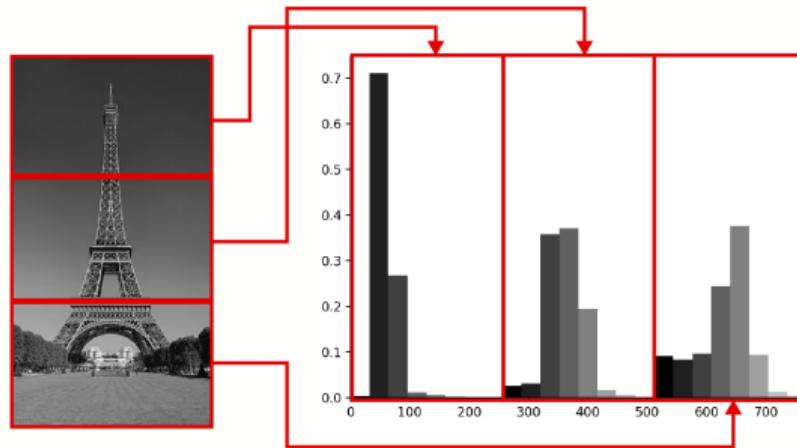
Description locale géométrique : description

Description de l'image :

1. Calcul d'un descripteur (hist. de couleurs, de LBP, etc.) par région
2. Concaténation des descripteurs

Contraintes :

- ▶ Découpage identique pour toutes les images
- ▶ Ordre de concaténation fixe pour un découpage donné
- ⇒ Géométrie du descripteur cohérente d'une image à une autre



Classification d'images :

- ▶ Avec histogrammes de couleurs global
- ▶ Avec histogrammes de LBP global
- ▶ Avec les deux
- ▶ Avec découpage géométrique des images