

Vision par ordinateur

Plan du chapitre

1. Introduction
2. Catégories de Computer Vision
3. Domaines d'utilisation
4. Processus de vision par ordinateur
5. Images



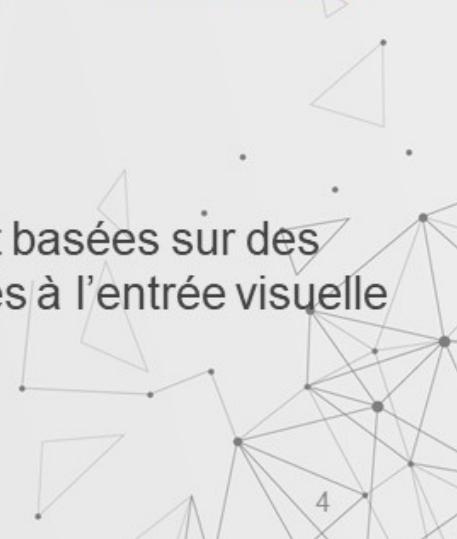
01

Introduction



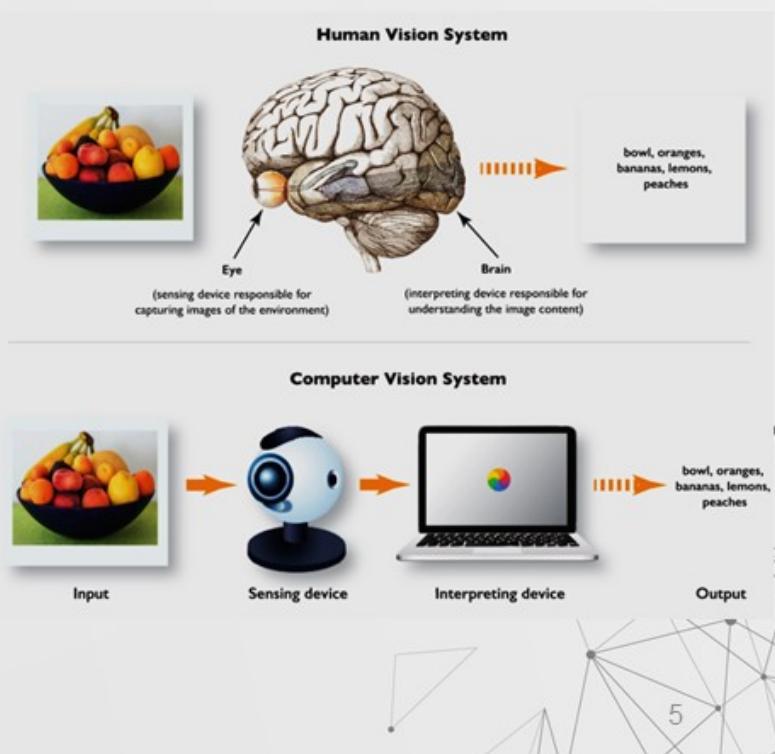
Introduction

- La vision par ordinateur désigne une technique d'intelligence artificielle permettant d'analyser des images saisies par un capteur tel qu'une caméra.
- Concrètement, la vision par ordinateur se présente comme un outil basé sur l'IA capable de:
 - reconnaître une image
 - de la comprendre
 - et de traiter les informations qui en découlent.
- La plupart des solutions de vision par ordinateur sont basées sur des modèles Machine Learning qui peuvent être appliqués à l'entrée visuelle des caméras, des vidéos ou des images.



Introduction

- La vision par ordinateur est l'équivalent, en termes d'IA, des yeux humains et de la capacité de notre cerveau à traiter et analyser les images perçues.
- La reproduction de la vision humaine par des ordinateurs constitue d'ailleurs l'un des grands objectifs de la vision par ordinateur.



Historique

1966	1970s	1980s	1990s	2000s	2010s
Camera attached to a Computer	First Image Processing Techniques	- Scale-space - Shape inference - Contour models	- Camera Calibration - 3D Reconstruction - Stereo Correspondence - Image Segmentation	- Autonomous vehicles - Facial recognition	- Deep neural networks - Convolutional neural networks (CNNs) - Object recognition and localization

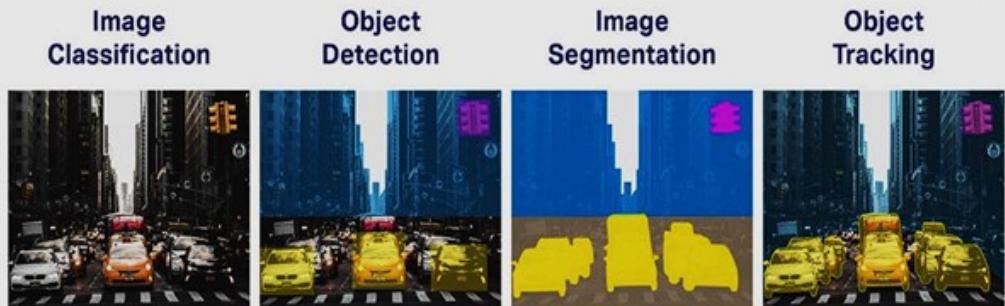
02

Catégories de Computer Vision

Les catégories de computer vision

- Le terme Computer Vision regroupe tous les algorithmes qui analysent et interprètent des images ou des vidéos.
- Il s'agit d'un domaine à part entière en intelligence artificielle, que l'on peut diviser en 4 grandes catégories :

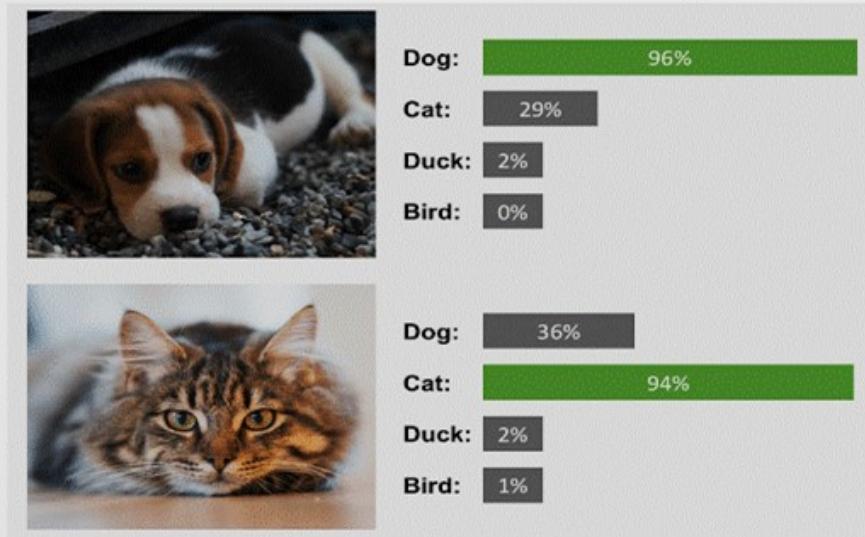
- **Classification**
- **Détection**
- **Identification**
- **Segmentation**



Classification des images

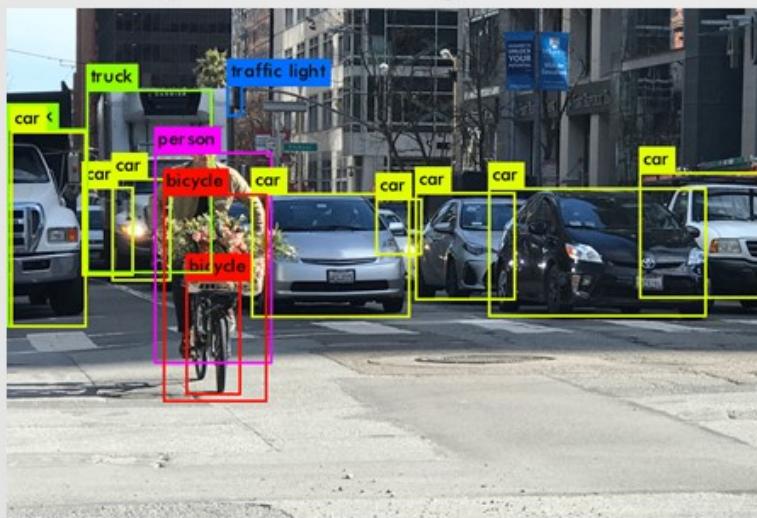
Un algorithme de classification prend en entrée une image à classer parmi une liste de possibilités.

Par exemple, un algorithme qui détermine à quelle espèce appartient un animal à partir de sa photo entre dans cette catégorie.



Détection

L'objectif est de détecter sur une image la position de certains objets. Les objets « à détecter » ont été définis à l'avance, il n'est pas nécessaire d'en détecter d'autres. La présence de ces objets est généralement représentée grossièrement par un rectangle et un nom sur l'image.



Identification d'image

L'identification est utilisée pour reconnaître un objet (souvent une personne). La différence avec les algorithmes précédents est que cet algorithme est capable de faire la différence entre 2 personnes différentes.

C'est ce genre d'algorithme qu'utilisent **Facebook**, **Apple** ou **Google Photos** pour nous identifier sur nos photos.



Segmentation

Ce type d'algorithme vise à délimiter les frontières entre les différents objets présents sur une image, et à classer ces derniers. Contrairement aux algorithmes de détection, ici chaque pixel appartient précisément à un unique objet.





03

Les domaines d'application

13

Voitures autonomes

Pour être capable de rouler sans conducteur, les voitures autonomes sont dotées de très nombreux capteurs, dont des caméras. Ces dernières filment l'environnement, notamment les panneaux de signalisation et le marquage au sol.

Les méthodes de Computer Vision sont alors utilisées pour interpréter ces images et permettre au véhicule d'agir en conséquence.



14

Reconnaissance faciale

Beaucoup de smartphones proposent aujourd’hui d’être déverrouillés par reconnaissance faciale. Les algorithmes utilisés varient en fonction des marques mais ils sont aujourd’hui très performants. De telles méthodes sont aussi utilisées par Facebook, Google Photos ou Snapchat pour reconnaître les visages sur les photos.



Imagerie médicale

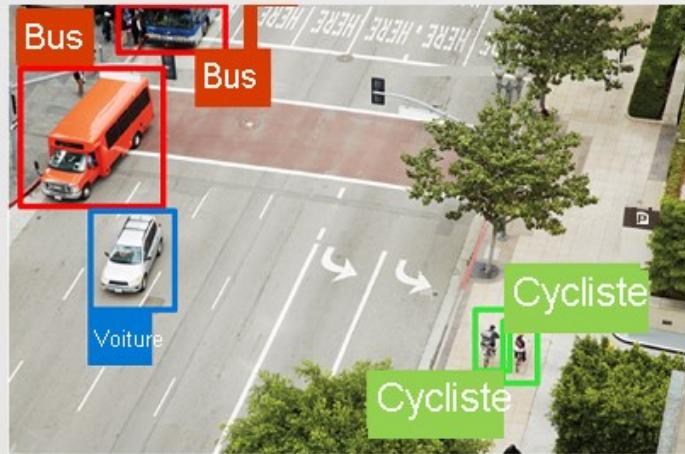
Sans s’être réellement imposée en tant que standard dans ce domaine, la Computer Vision montre des résultats plus que prometteurs dans la détection de maladie. Une équipe de chercheurs du MIT a notamment développé une IA rivalisant avec les médecins pour détecter précocement les cancers du sein en utilisant les mammographies des patients.



Supervision du trafic

Une solution de supervision du trafic peut utiliser:

- la détection d'objet pour identifier l'emplacement de différentes classes de véhicules
- la classification d'images afin de classifier les images en fonction du type de véhicule qu'elles contiennent, comme les taxis, les bus, les cyclistes, etc.



Segmentation sémantique

La segmentation sémantique est une technique avancée de Machine Learning dans laquelle les pixels individuels de l'image sont classifiés en fonction de l'objet auquel ils appartiennent.

Par exemple, une solution de supervision du trafic peut superposer des images de trafic avec des couches de « masques » pour mettre en évidence différents véhicules à l'aide de couleurs spécifiques



Analyse d'image

On peut créer des solutions qui combinent des modèles Machine Learning avec des techniques avancées d'analyse d'image afin d'extraire des informations à partir d'images, y compris des « étiquettes » qui permettent de cataloguer l'image, ou même des légendes descriptives qui résument la scène illustrée dans l'image



Détection, analyse et reconnaissance des visages

La détection des visages est une forme spécialisée de détection des objets qui localise les visages humains dans une image. Elle peut être combinée avec des techniques de classification et d'analyse de la géométrie faciale pour reconnaître des individus en fonction de leurs traits



Reconnaissance optique de caractères

La reconnaissance optique de caractères (OCR) est une technique utilisée pour détecter et lire du texte dans des images. On peut utiliser la reconnaissance optique de caractères pour lire du texte sur des photos (par exemple, des panneaux de signalisation routière ou des vitrines de magasins), ou pour extraire des informations de documents numérisés tels que des lettres, des factures ou des formulaires



04

Processus de Vision par ordinateur



21

22

Système de vision par ordinateur

Connaissances et technologies qui permettent de concevoir des machines qui peuvent « voir »

□ Premier niveau : vision

- acquiert une image grâce un capteur

□ Deuxième niveau (bas niveau) : traitement de l'image

- modifie le contenu de l'image afin de mettre en évidence des éléments d'intérêt (objets, contours)

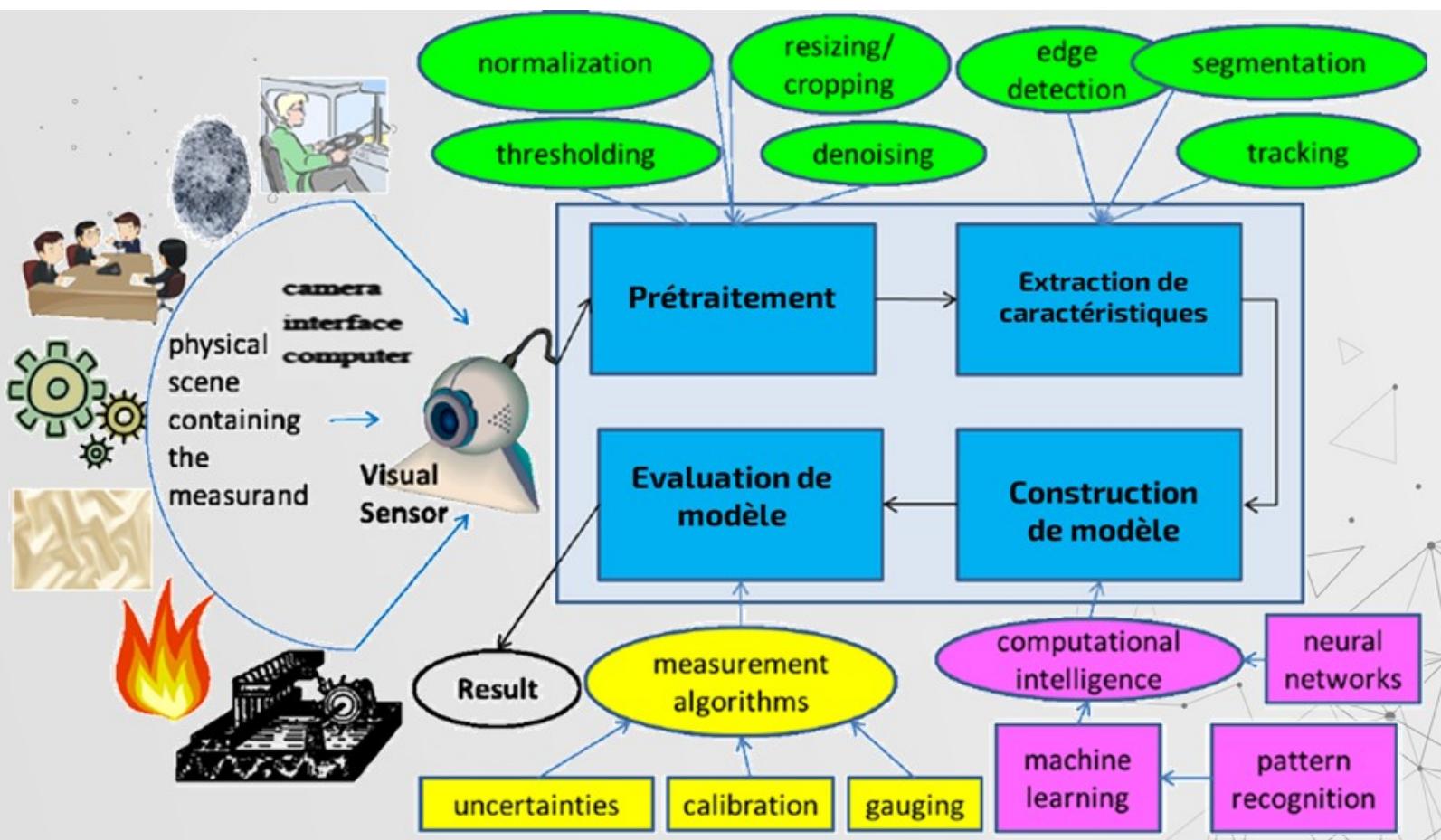
□ Troisième niveau (Haut Niveau) : reconnaissance

- utilise des techniques d'intelligence artificielle pour identifier des formes connues dans l'image

Processus de vision par ordinateur

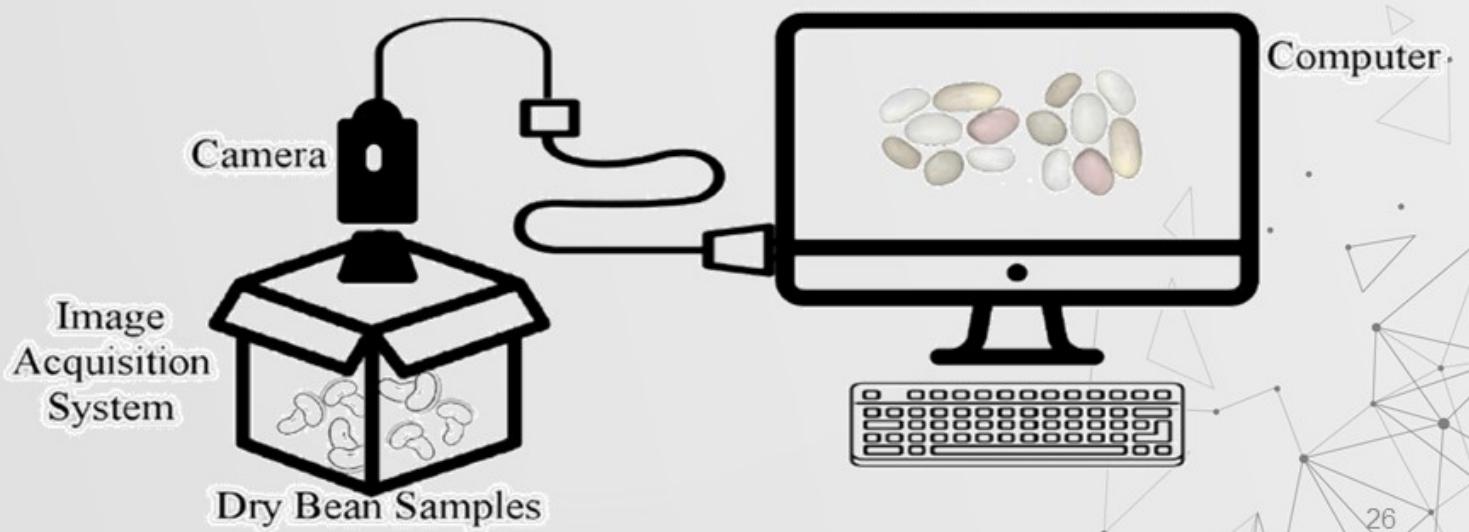
Un processus de vision par ordinateur est une série d'étapes de traitement appliquées aux images ou aux vidéos d'entrée pour extraire des informations pertinentes ou prendre des décisions basées sur des données visuelles.

Voici un processus généralisé pour les tâches de vision par ordinateur :



Etape 1: Acquisition

Des images ou des séquences vidéo sont acquises par ordinateur, même en temps réel, en utilisant des photos ou la technologie 3D, la vidéo, à des fins d'analyse.



Etape 2: Prétraitement

Une fois les images acquises, elles doivent être prétraitées. Cela implique des tâches telles que l'amélioration de l'image, la suppression du bruit, la segmentation de l'image, etc.

- **Redimensionner les images**: standardisez la taille des images d'entrée.
- **Normalisation**: normalisez les valeurs des pixels selon une plage standard (par exemple, 0 à 1).
- **Conversion de l'espace colorimétrique**: convertissez les images dans un espace colorimétrique approprié (par exemple, RVB, niveaux de gris).
- **Réduction du bruit**: appliquez des filtres ou des techniques pour réduire le bruit dans les images.
- **Augmentation des données**: introduisez des variations dans les données d'entrée pour améliorer la généralisation du modèle.

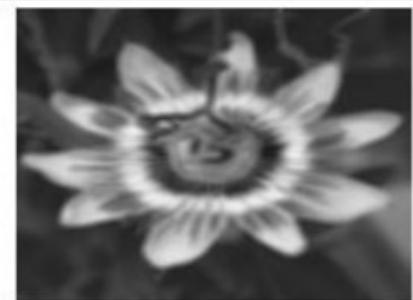
Exemple de prétraitement des images



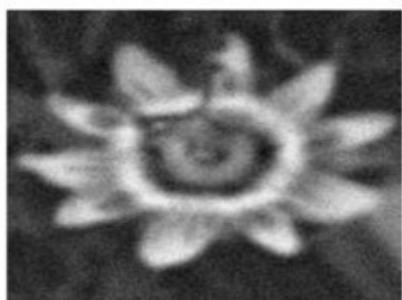
(a) Image originale



(b) Bruit gaussien, moyenne nulle, écart-type $\sigma = 50$



(c) Flou



(d) Bruit et flou



(e) Bruit poivre et sel



(f) Bruit multiplicatif

Etape 3: Extraction de caractéristiques

- **L'extraction de caractéristiques visuelles** (ou visual features extraction en anglais) consiste en des transformations mathématiques calculées sur les pixels d'une image numérique.
 - Les caractéristiques visuelles permettent généralement de mieux rendre compte de certaines propriétés visuelles de l'image, utilisées pour des traitements ultérieurs entrant dans le cadre d'applications telles que la détection d'objets ou la recherche d'images par le contenu.
 - Large variété de représentations suivant les caractéristiques utilisées
- | | |
|---|---|
| Caractéristiques globales: | Caractéristiques locales: |
| <ul style="list-style-type: none">• Surface, périmètre, largeur, hauteur• Elongation, circularité• Moments statistiques | <ul style="list-style-type: none">• Coins ou sommets (nombre, positions relatives ou absolues, angles, ...)• Segments (nombre, positions relatives ou absolues, longueur, ...) |

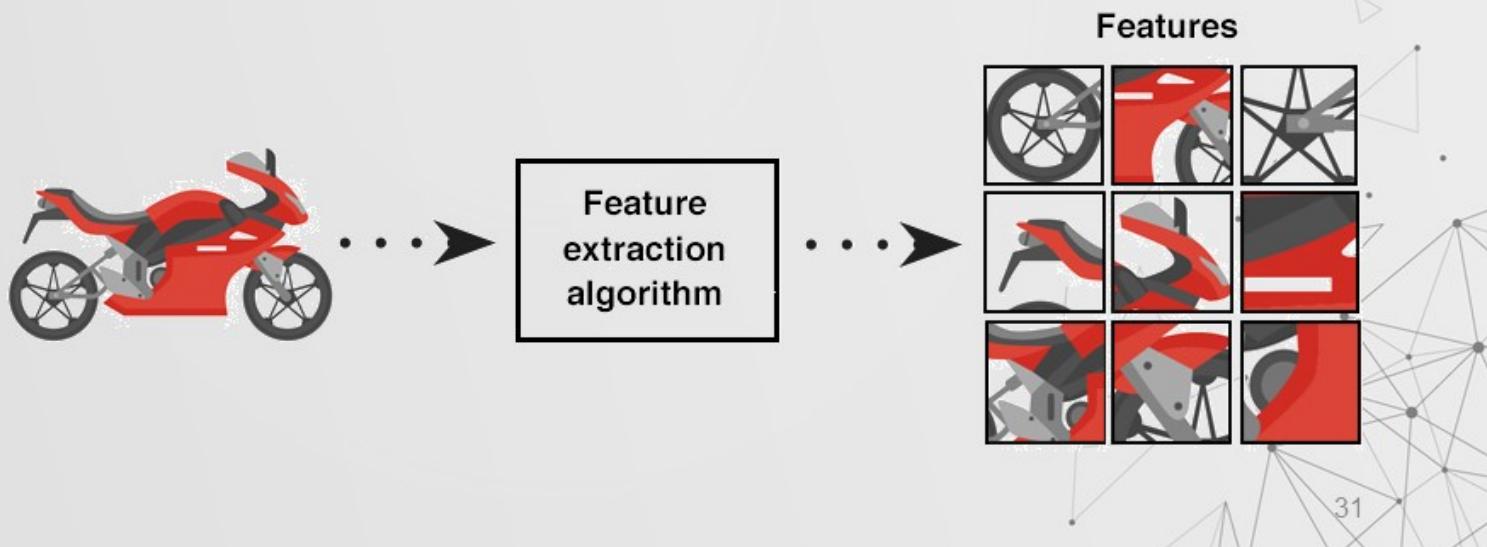
Etape 3: Extraction de caractéristiques

- Les caractéristiques peuvent être extraites sur:
 - la forme elle-même
 - le squelette
 - les contours
- Les points caractéristiques se définissent, en général, à partir des contours. Ils comprennent les **extrémités** de lignes et **les points de forte courbure**.
- Ces points sont exploités pour la fermeture de contours et l'élaboration de contours virtuels. Ils constituent aussi des points de focalisation de l'attention en interprétation d'images.

Etape 3: Extraction de caractéristiques

Les caractéristiques sont les caractéristiques des objets dans les images qui sont extraites et utilisées pour une analyse plus approfondie.

Les caractéristiques peuvent être la couleur, la taille, la forme, la texture, etc.



Etape 4: Construction de modèles

La vision par ordinateur est un domaine qui englobe de nombreux algorithmes pour traiter et interpréter des données visuelles. Voici quelques-uns des algorithmes couramment utilisés dans la vision par ordinateur :

1. Réseaux de neurones convolutionnels (CNN) :

Principalement utilisés pour la classification d'images, la détection d'objets et la segmentation sémantique.

2. Réseaux de neurones récurrents (RNN) :

Utilisés pour le traitement séquentiel, par exemple, dans la reconnaissance d'actions dans des vidéos.

3. Algorithmes de détection d'objets :

YOLO (You Only Look Once), Faster R-CNN, SSD (Single Shot MultiBox Detector), etc.

4. Réseaux de neurones pré-entraînés :

Utilisation de modèles pré-entraînés comme VGG, ResNet, Inception, etc., pour diverses tâches de vision par ordinateur.

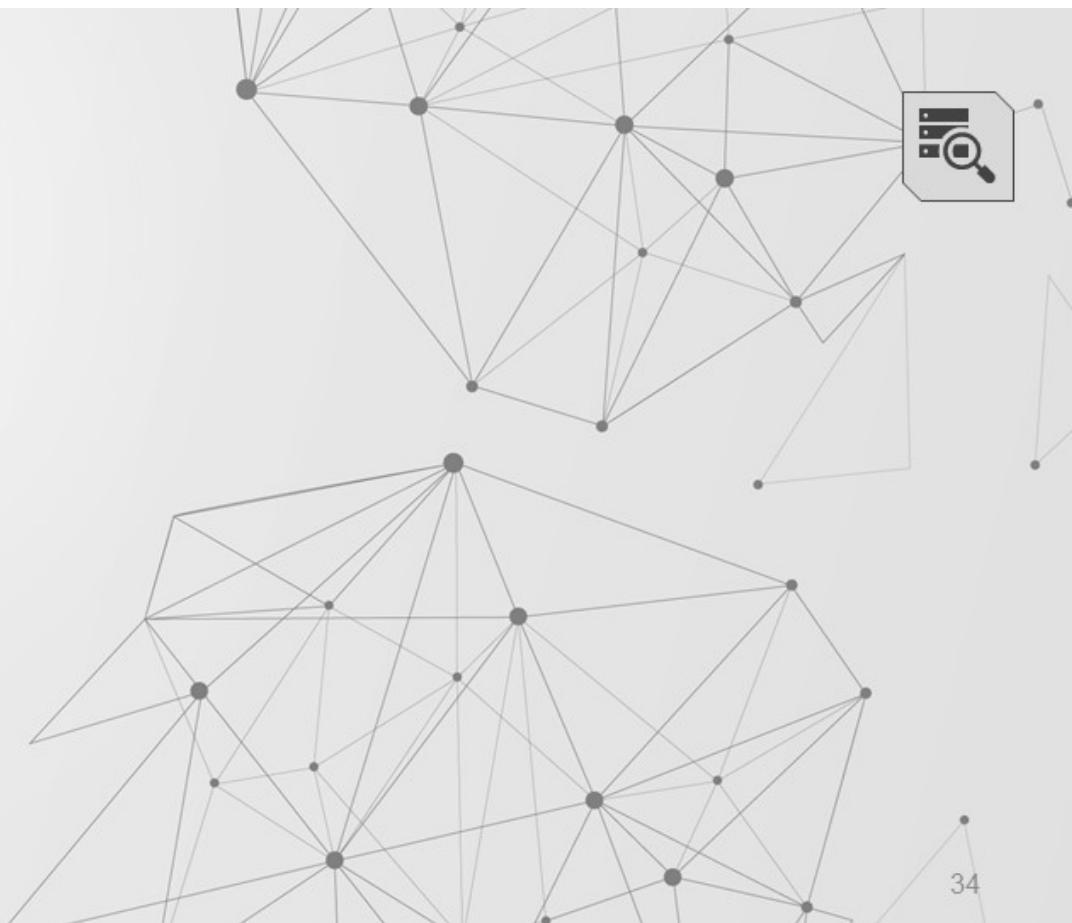
Etape 5: Evaluation de modèles

L'évaluation des modèles de vision par ordinateur est cruciale pour mesurer leur performance et comprendre leur efficacité dans la résolution de tâches spécifiques. Voici quelques méthodes couramment utilisées pour évaluer les modèles de computer vision :

- **Précision (Accuracy)**
- **Précision (Precision) et Rappel (Recall)**
- **F1-score**
- **Matrice de confusion**
- **Etc.**

Il est souvent recommandé d'utiliser plusieurs métriques en fonction de la nature spécifique de la tâche de vision par ordinateur que vous effectuez. Certaines métriques peuvent être plus pertinentes pour certaines tâches, et le choix dépend des exigences spécifiques de votre problème.

05 Images



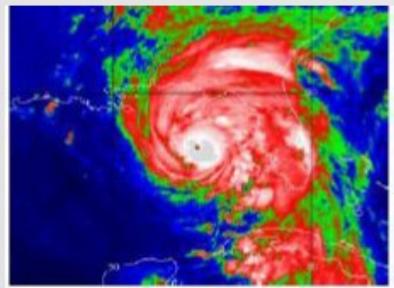
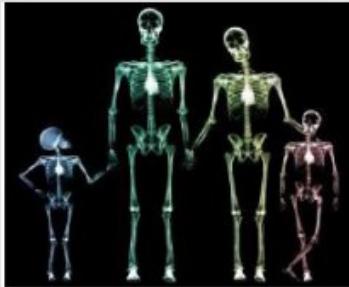
Différents types d'images

- **Images naturelles**

Moyens d'acquisition: Caméras, Scanners, satellites, infra-rouge, IRM....

- **Images artificielles**

Outils de représentation: synthèse d'images, réalité virtuelle....



Définition d'une image

Une image est une forme discrète d'un phénomène continu obtenue après discréttisation (numérisation)

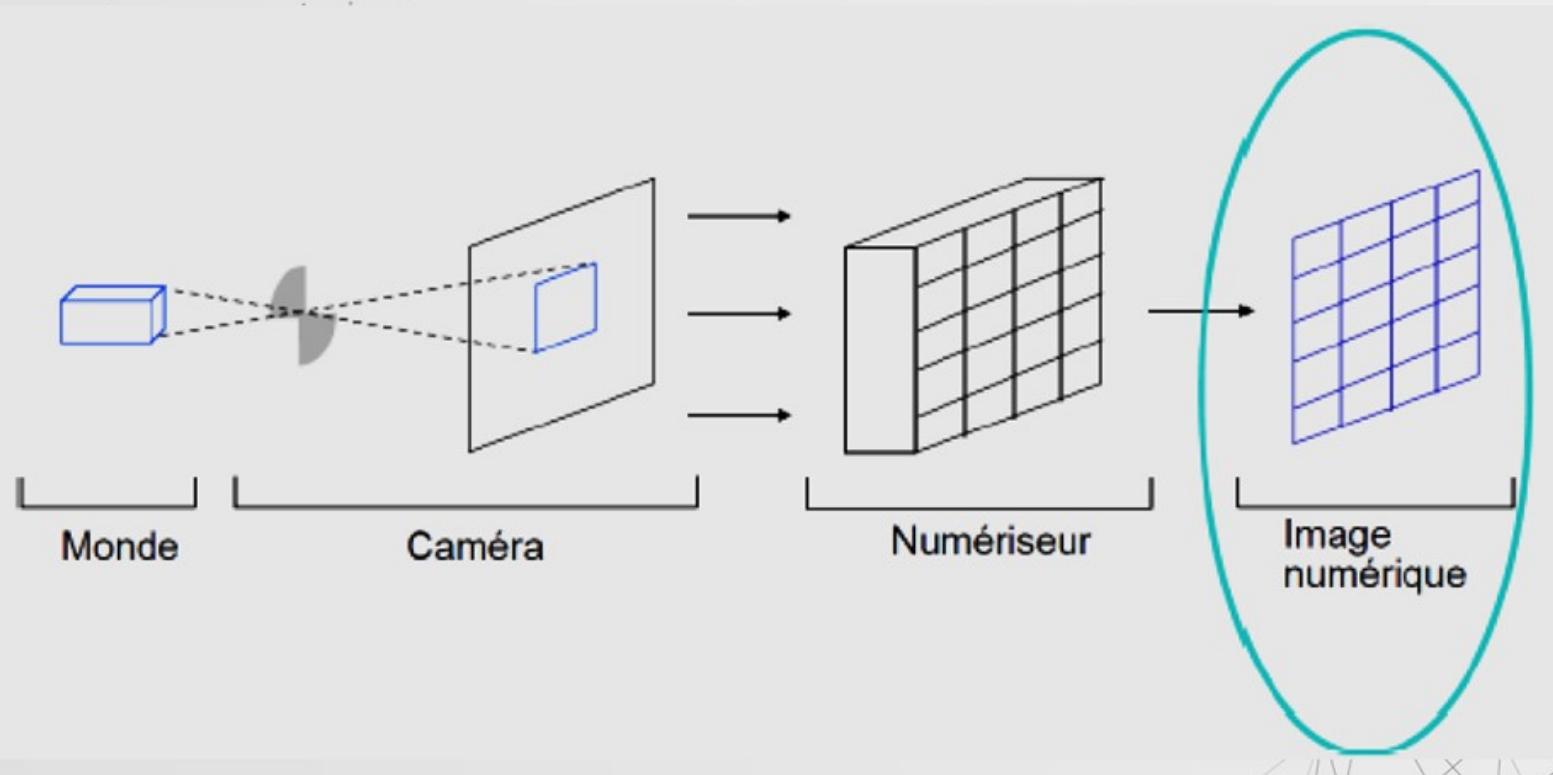
Une image est une fonction $f: \mathbb{R}^2 \rightarrow \mathbb{R}$

$F(x,y)$ est l'intensité en (x,y)

Image Couleur= trois fonctions

$$f(x, y) = \begin{bmatrix} r(x, y) \\ g(x, y) \\ b(x, y) \end{bmatrix}$$

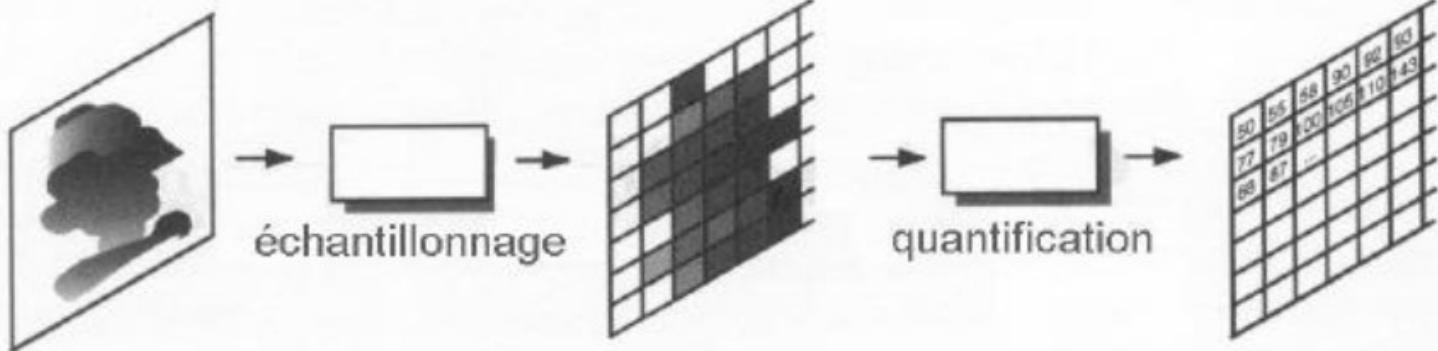
Acquisition de l'image



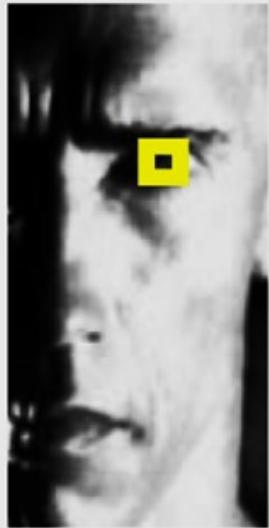
Numérisation

En vision par ordinateur, nous travaillons généralement sur des images numériques (discrètes).

Numérisation = Échantillonnage + Quantification



Numérisation



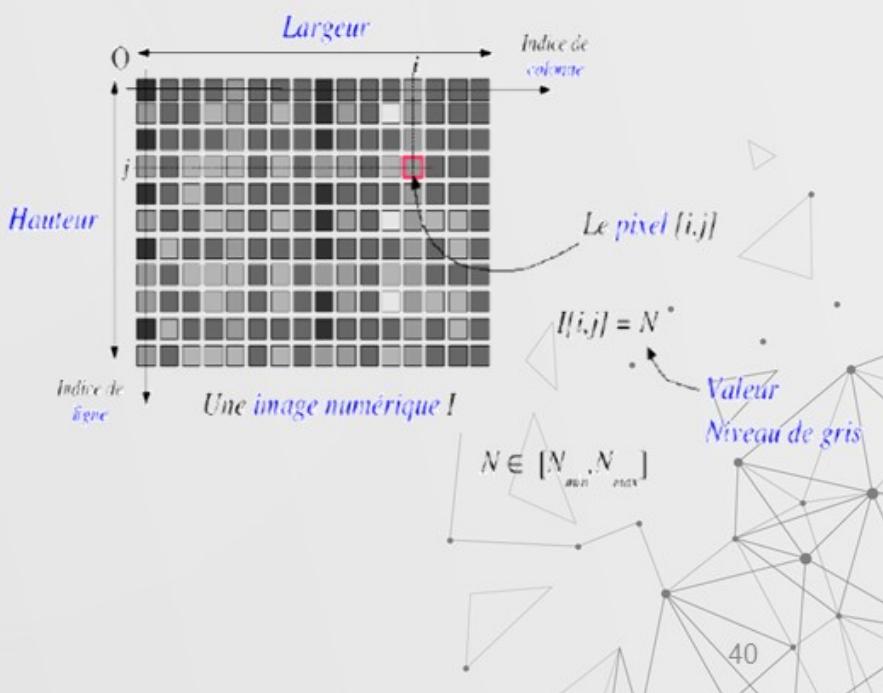
A coordinate system with x and y axes is shown next to the image. A yellow square highlights a specific pixel on the forehead.

62	79	23	119	120	105	4	0
10	10	9	62	12	78	34	0
10	58	197	46	46	0	0	48
176	135	5	188	191	68	0	49
2	1	1	29	26	37	0	77
0	89	144	147	187	102	62	208
255	252	0	166	123	62	0	31
166	63	127	17	1	0	99	30

Représentation d'une image

Une **image** numérique est représentée par une matrice I ($N \times N$),

- Chaque élément de la matrice représente l'intensité d'un pixel (Picture Element)
- Un **Pixel** est l'unité de base permettant de mesurer la définition d'une image
- Une **image** est caractérisée par ses dimensions en nombre de pixels et non en centimètres.



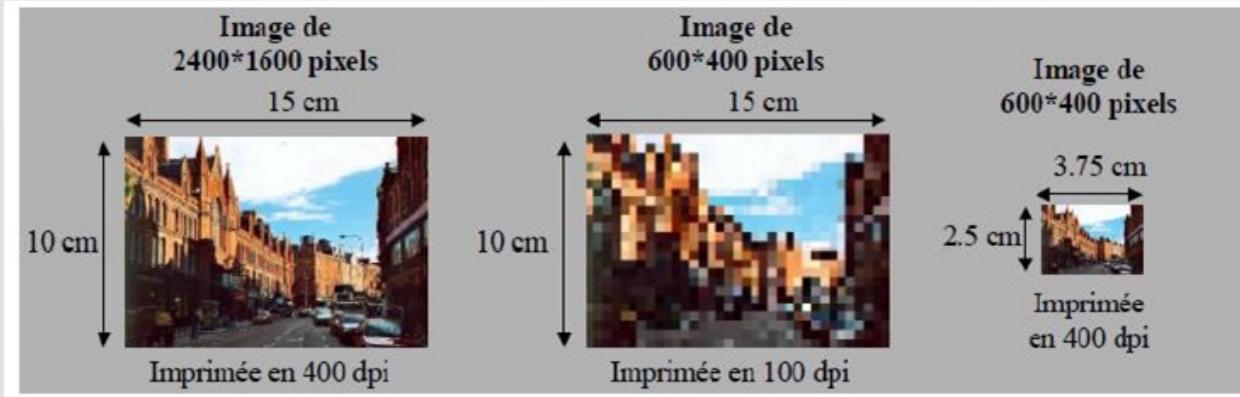
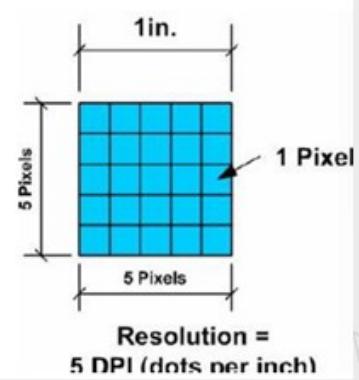
Résolution

Résolution : le nombre de pixels par unité de longueur

On peut calculer une résolution en

- Pixel par pouce (ppp)
- dot per inch (dpi)

(1 inch = 1 pouce = 2.54 cm)



Résolution- Exemple

Résolution...

...spatiale :



256x256



128x128



64x64



32x32

Échantillonnage

...tonale :

Quantification



6 bits



4 bits



3 bits



2 bits



1 bit

Différents types d'image

- **Image binaire Noir et blanc** codé en $(0,1)$ $m=1 \in \{0,1\}$
- **Image en Niveau de Gris** : $2^8 = 255$ nuances $m=8; k \in \{0,255\}$

par convention : noir = 0 et blanc = 255 (ou 1 si normalisation)

Images couleurs (3 Octets) $m=3*8 =24$ bits

$K \in \{0, \dots, 16777255\}$

01100010
01101001
01101110
01100001
01110010
01111001

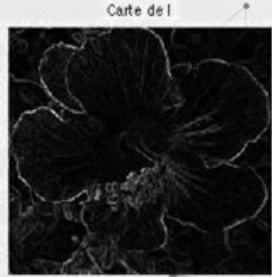
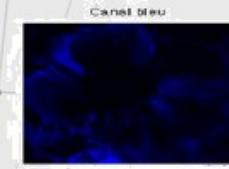
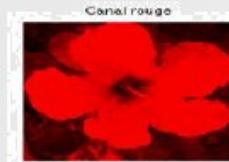
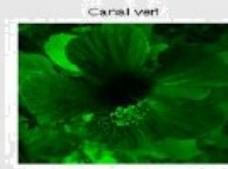


Image Couleur

Image en couleur indexée dans une palette (table) de couleurs

- L'espace couleur est basé sur la synthèse additive des couleurs,
- Mélange de trois composantes (par exemple (R, V, B)) donne une couleur.
- Un pixel est codé par trois valeurs numériques...

R				0				255
V				0				255
B				255				255
index	6	255



Contenu d'une image

- **Texture** : répartition statistique ou géométrique des intensités dans l'image
- **Contour** : limite entre deux (ou un groupe de) pixels dont la différence de niveaux de gris (ou de couleurs) est significative
- **Région** : groupe de pixels présentant des caractéristiques similaires (intensité, mouvement, etc.)
- **Objet** : région (groupe de régions) entièrement délimitée par un contour, possédant une indépendance dans l'image
- **description sémantique** : on peut donner un nom à un objet



Formats d'images numériques

Données brutes : raw data

- **Standards universels** : gif, bitmap, tiff, ppm, eps, ...
- **Standards médicaux** : DICOM

Quelques exemples :

- **BMP (Bitmap)** : matrice de bits codés en couleur (jusqu'à 24bits/pixel)
- **GIF** : format compressé avec codage 8 bits/pixel
- **JPG (jpeg)** : format de compression d'images photographiques