



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



La vision par ordinateur

Chapitre 1 : Introduction à la vision par ordinateur

Master 2 : Systèmes Informatique Intelligents
Lyes_sii@yahoo.fr lyes_abada@yahoo.fr

- 1- Introduction à la vision par ordinateur
- 2- Techniques de base de traitement de l'image
- 3- Géométrie, Calibration et Mathématique Projective
- 4- La reconstruction 3D (géométrique et photométrique)
- 5- La texture en vision par ordinateur
- 6- Le mouvement en vision par ordinateur
- 7- La vision pour la Bio-Informatique et la Biométrie

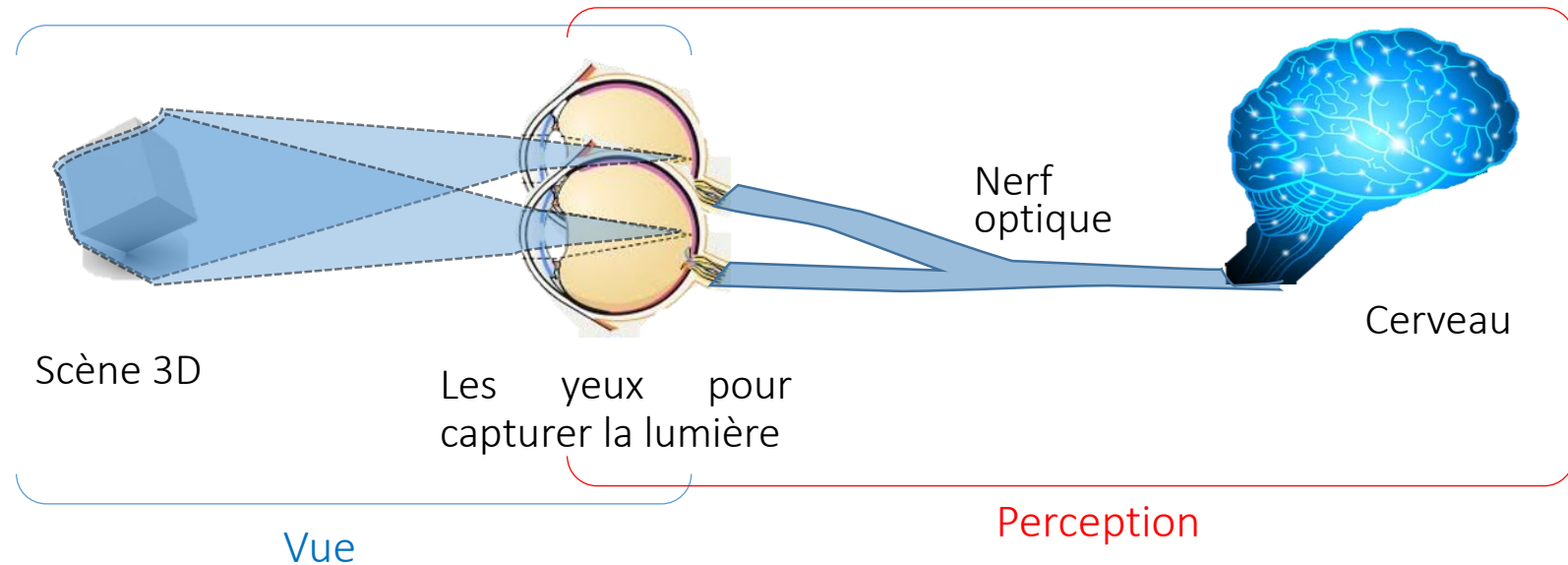
Introduction

La vision humaine est **un système compliqué** de la perception des objets distants.

Les neurophysiologistes mènent des recherches théoriques et expérimentales **afin de comprendre l'anatomie et le fonctionnement du cerveau** dans son ensemble ;

ils ont découvert une structure **très complexe** qui est loin de leur révéler tous ses secrets.

La vision humaine **décrit le monde qui l'entoure aisément**. Il s'agit d'un système complexe qui se compose de deux organes qui sont : **l'œil et le cerveau**.



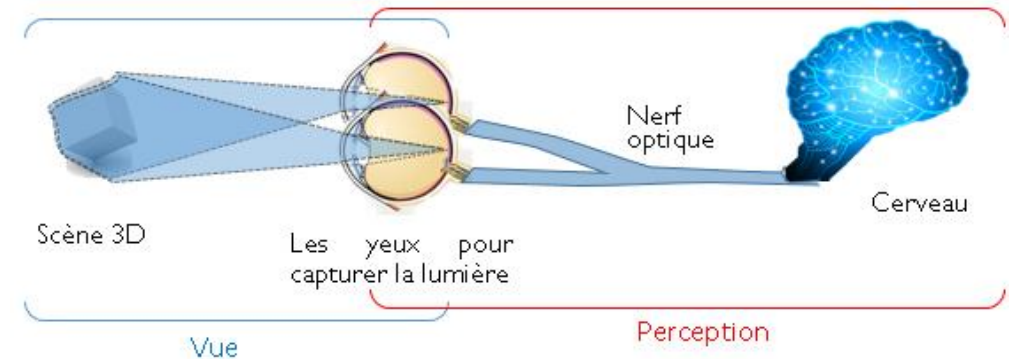
Vision humaine

L'œil de l'être humain reçoit **une quantité très importante d'informations**.
Il est très sensible aux spectres des ondes électromagnétiques (**lumière visible**).
Ces organes fournissent **une vision stéréoscopique** (c'est-à-dire 3D).

Les rayons lumineux d'ordre **d'un million de points** environ sont reçus par la rétine, ils contiennent des informations sur **la quantité de lumière** et de **couleurs** provenant de l'environnement.

Ils forment **l'image** qui sera conduite sous forme de signaux transmissibles le long des voies neuronales vers le cerveau.

A ce stade, le cerveau prend le relais et donnera une signification à l'image perçue par l'œil.



Vision humaine



Nous avons donc pu capturer la lumière et la couleur comme l'œil humain,

Comprendre ce qu'il y a sur les photos est l'opération la plus difficile pour une machine.



<https://fr.freepik.com/photos/fleur>

Si on prend cette photo, nos cerveaux peut facilement savoir que c'est une fleur.

95	48	122	153	193	193	18	173	88	163	109	53	44		
106	87	166	240	205	155	88	201	238	181	47	44	41		
169	79	236	216	219	15	14	7	31	23	125	71	23		
248	287	13	196	9	113	3	126	186	37	219	3	55		
157	215	67	73	64	38	218	135	52	86	88	252	215		
253	14	174	78	225	225	189	156	99	47	225	212	85		
76	171	125	23	68	84	28	109	184	49	13	238	185		
51	33	258	237	74	174	53	241	94	18	61	237	188		
165	156	97	42	232	255	189	241	283	34	23	218	233		
125	219	132	231	145	64	283	135	3	87	124	37	94		
163	4	236	58	233	244	176	66	139	164	18	11	106		
72	198	223	133	78	238	24	235	28	42	127	223	167		
158	226	163	114	121	1	194	245	215	199	196	212	242		
181	135	194	45	192	4	16	74	12	185	188	128	178		
88	114	247	18	149	215	141	128	148	241	218	23	23		
59	137	120	190	3	124	24	18	20						
194	17	87	115	178	118	89	59	255	214	182	48	191		
212	11	215	184	238	132	118	188	179	221	99	76	3		

Mais ce n'est pas le cas pour la machine. Dont l'image est représentée par un tableau de valeurs entières qui représentent les intensités et les couleurs.

Il n'y a pas de contexte, Juste une grande pile de données!!!

Vision par ordinateur

Aujourd'hui, l'être humain tente de reproduire la vision humaine par un système similaire. Ceci a donné naissance à ce qu'on appelle « **vision par ordinateur** », « **vision artificielle** » ou « **vision cognitive** ». Celle ci représente une discipline de l'intelligence artificielle.

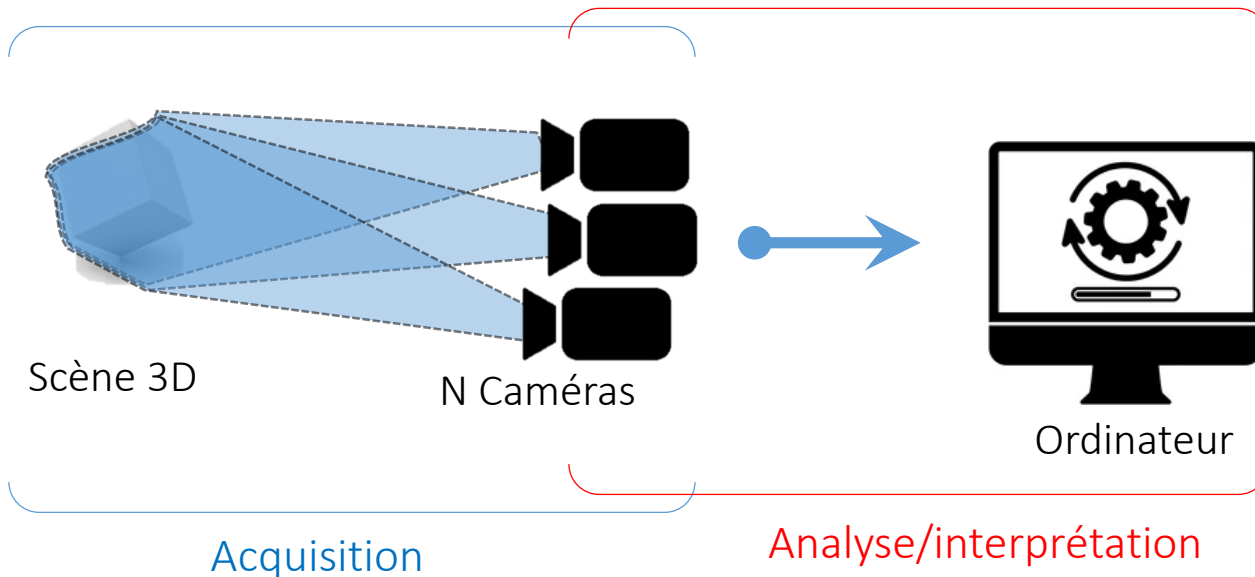
La vision par ordinateur (vision Artificielle) est une branche de l'Informatique dont le but est de rendre une machine capable de « **comprendre** » une scène donnée, en se basant sur des informations essentiellement visuelles.

La vision par ordinateur est constituée de l'ensemble des technologies de traitement de l'information numérique ayant pour origine des capteurs photométriques

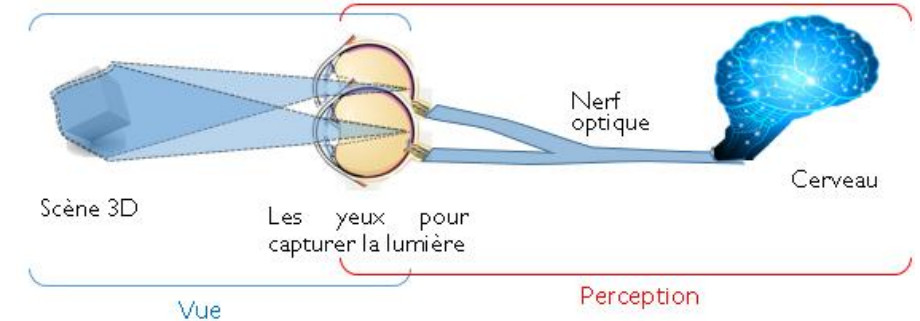
Vision par ordinateur

Un système de vision artificielle prend en entrée une ou plusieurs images numériques et effectue des traitements sur ces images afin d'en extraire des informations relatives à la scène observée

Contrairement à la perception humaine qui est basée sur deux images (œil droit et œil gauche), l'ordinateur peut quant à lui utiliser des sources d'informations plus nombreuses et plus variées, éventuellement dans d'autres bandes spectrales que le visible

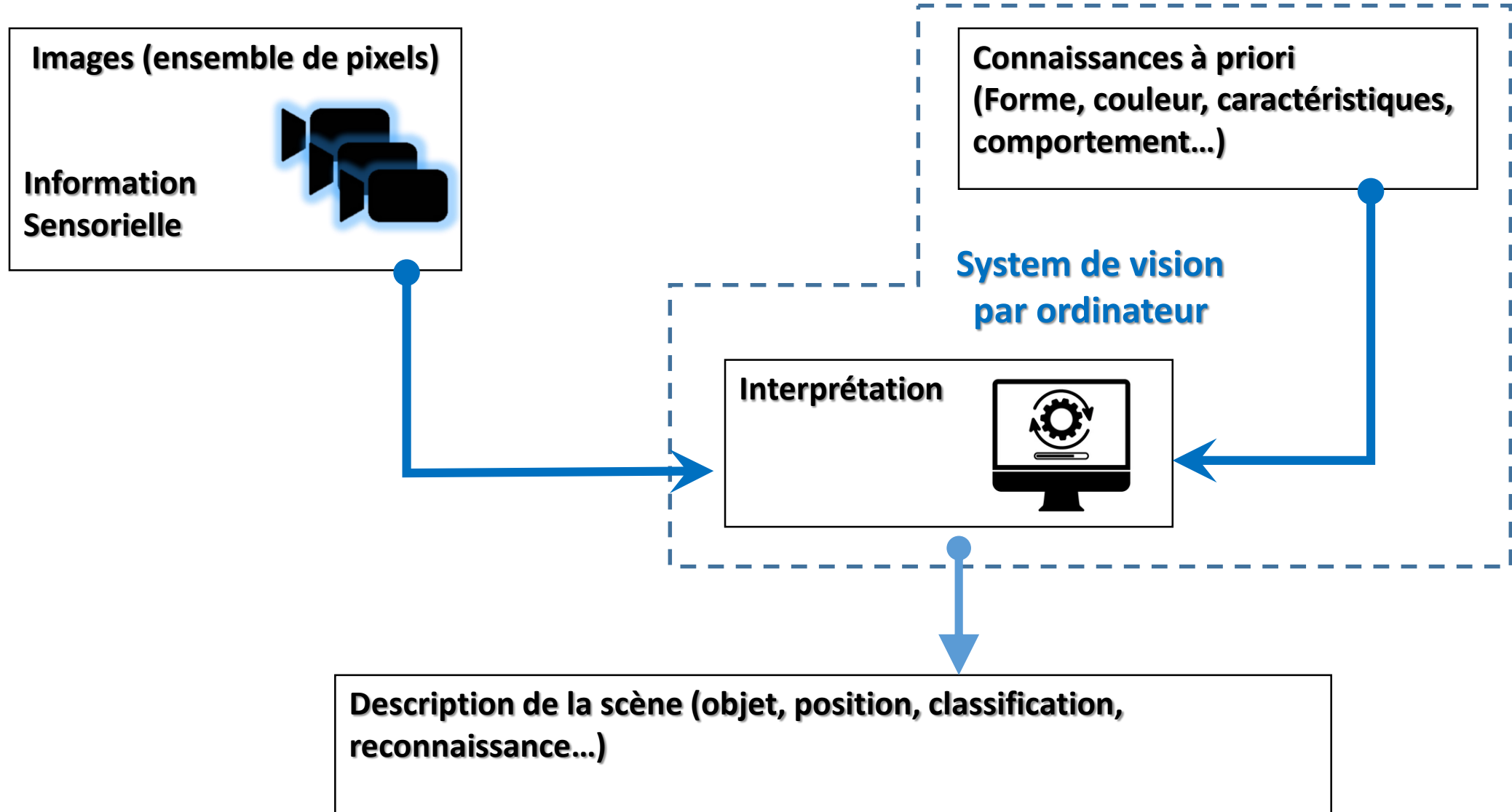


Vision artificielle



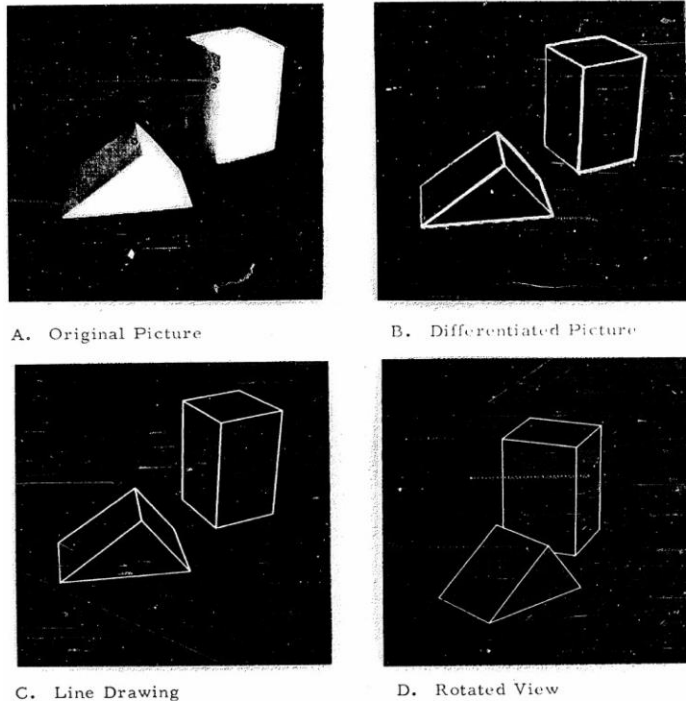
Vision humaine

Fonctionnement d'un système de Vision



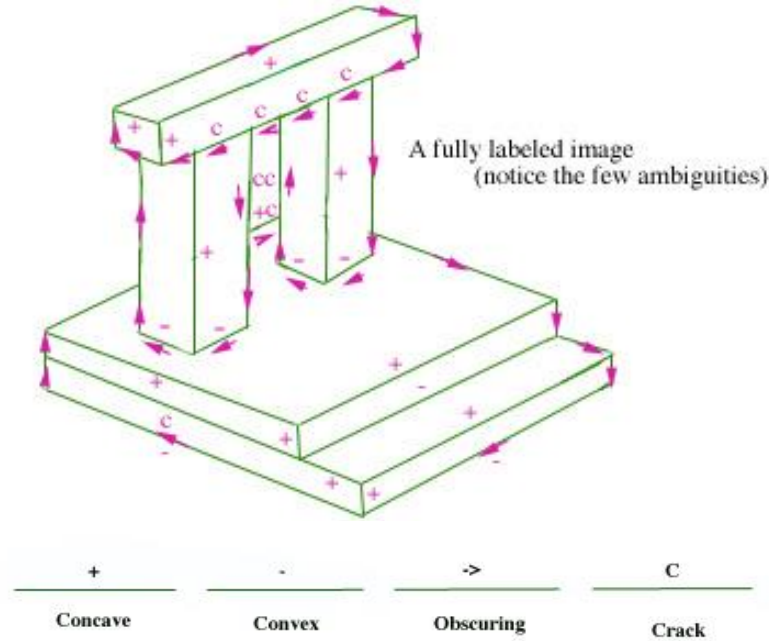
Histoire de la vision par ordinateur

- 1970s
 - Récupérer la structure 3D du monde à partir d'images

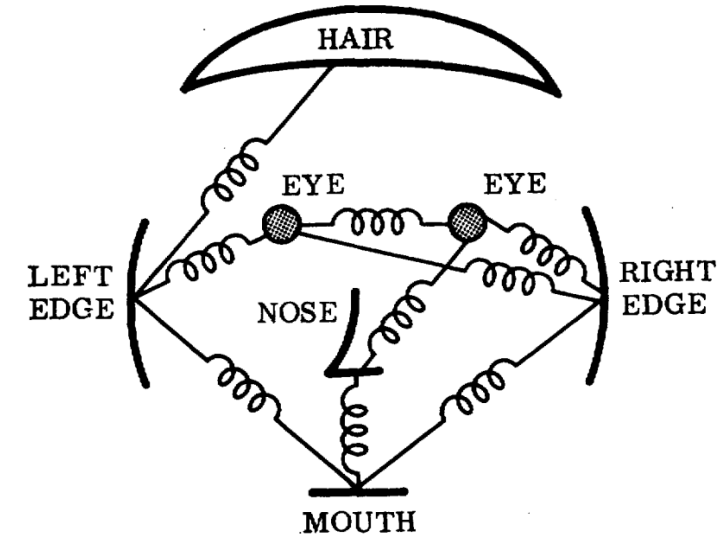


Blocks World

Roberts: Machine perception of three-dimensional solids. PhD Thesis, 1963



Line Labeling



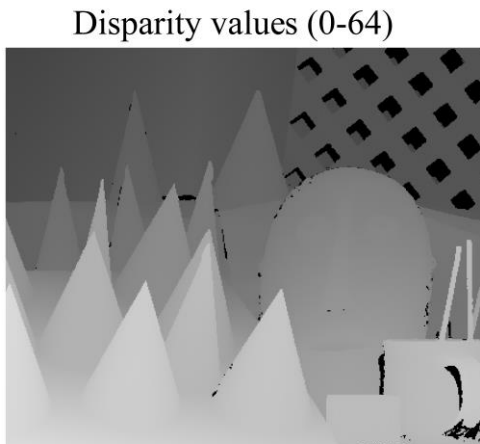
Pictorial Structure
Fischler and Elschlager 1973

Histoire de la vision par ordinateur

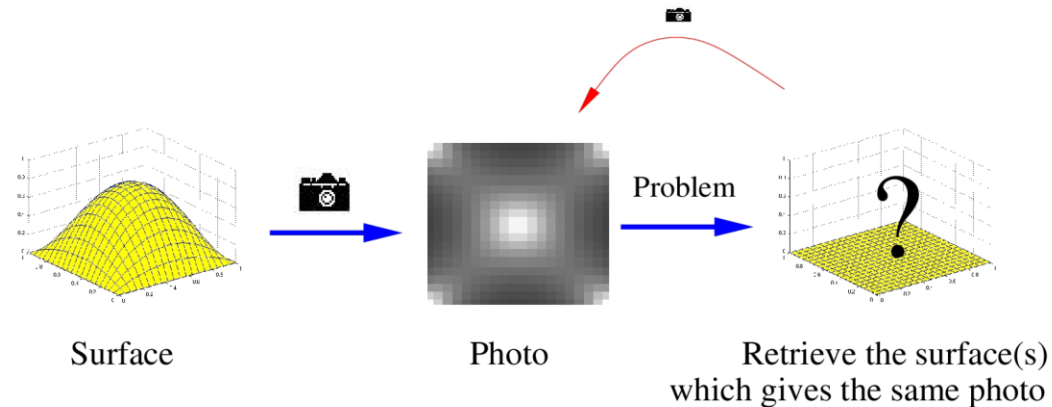
- 1980s
 - Algorithmes de correspondance stéréo et algorithmes de flux optique.
 - Shape from X techniques (shape from shading, shape from texture, shape from shadows)
 - Détection des bords et des contours



Canny edge detector. Canny, 1986



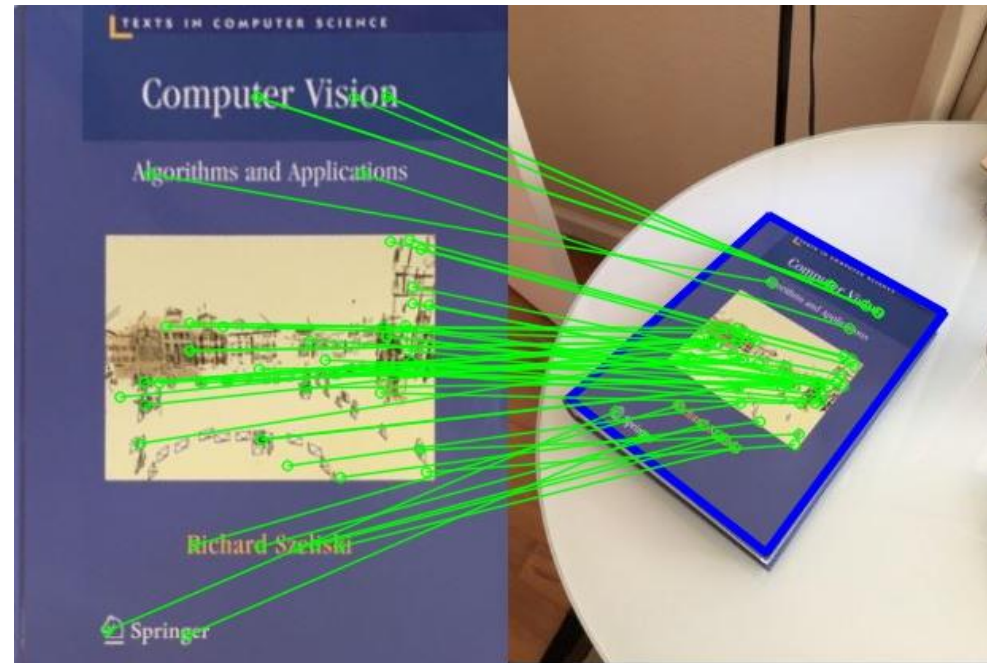
Note how disparity is larger (brighter) for closer surfaces.



Shape from shading

Histoire de la vision par ordinateur

- 1990s
 - Structure from Motion and Multi-view Reconstruction
 - Scale Invariance Feature Transform (SIFT)



David Lowe: Object recognition from local scale-invariant features. ICCV, 1999.

Histoire de la vision par ordinateur

- 1990s
 - Les techniques d'apprentissage statistique ont commencé à apparaître.

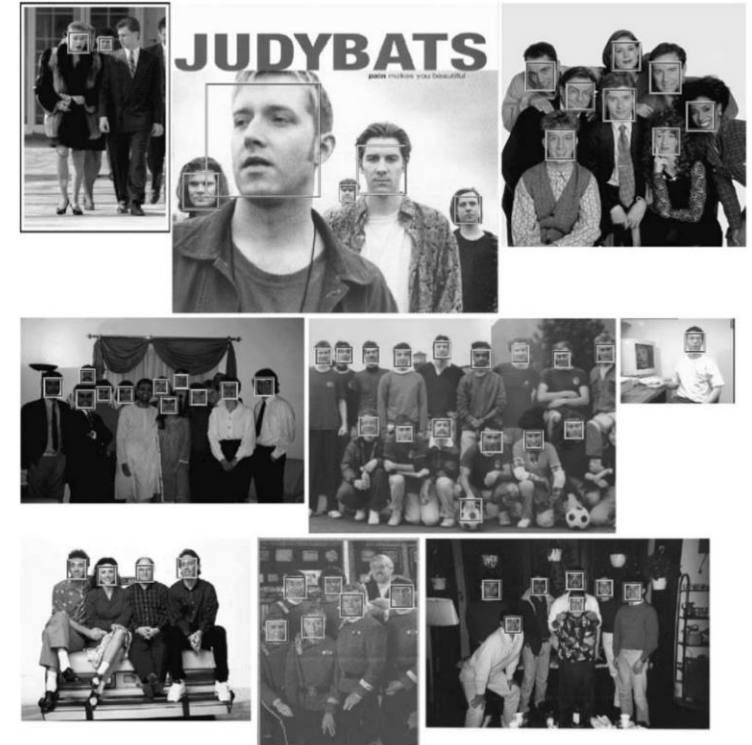


Eigenfaces

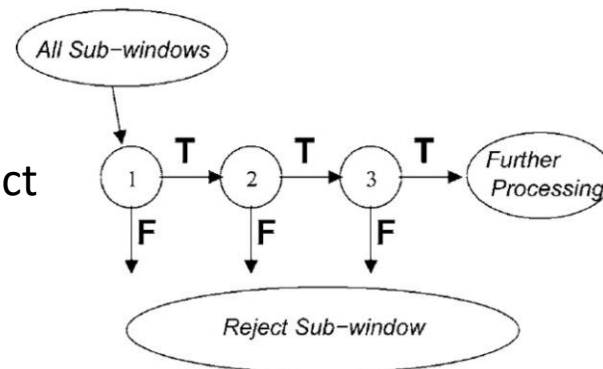
Turk and Pentland: Face recognition using Eigenfaces. CVPR, 1991

Histoire de la vision par ordinateur

- 2000s
 - Approches basées sur les données et l'apprentissage (Data-driven and learning approaches)
 - Classificateurs en cascade (Cascaded classifiers) pour la détection d'objets



Viola and Jones: Robust Real-time Object Detection. IJCV, 2001.



Histoire de la vision par ordinateur

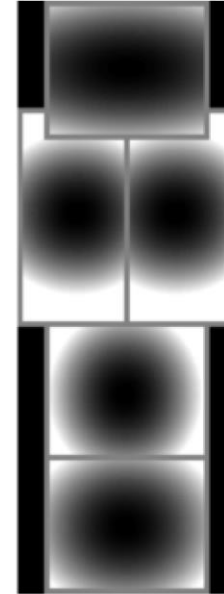
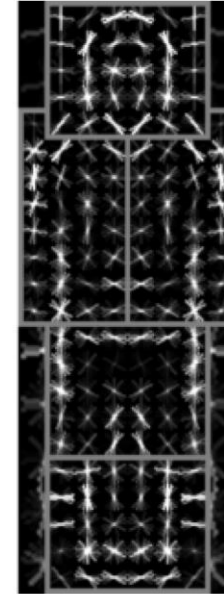
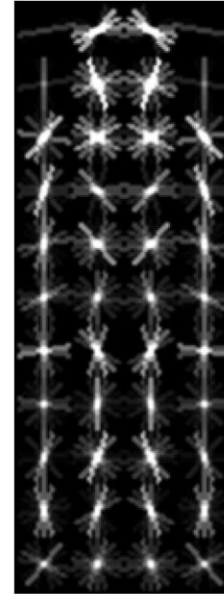
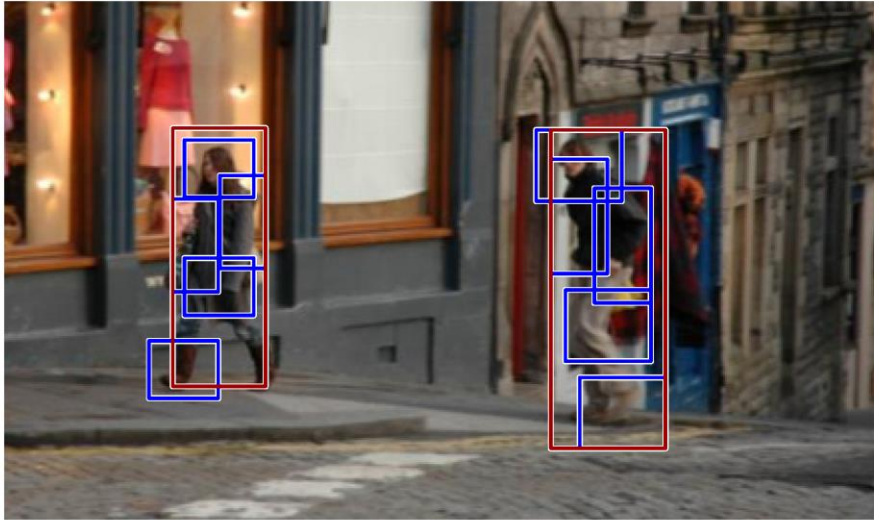
- 2000s
 - Histogramme des gradients orientés pour la détection d'objets



Dalal and Triggs: Histograms of Oriented Gradients for Human Detection. CVPR, 2005.

Histoire de la vision par ordinateur

- 2000s
 - Deformable parts models for object detection



Felzenszwalb et al. Object detection with discriminatively trained part-based models . TPAMI, 2009.

Histoire de la vision par ordinateur

- 2000s
 - Datasets



The PASCAL Visual Object Classes Challenge 2007



PASCAL VOC, Everingham et al., 2005 - 2012



ImageNet, Deng et al., 2009

Histoire de la vision par ordinateur

- 2000s
 - Large-scale structure from motion

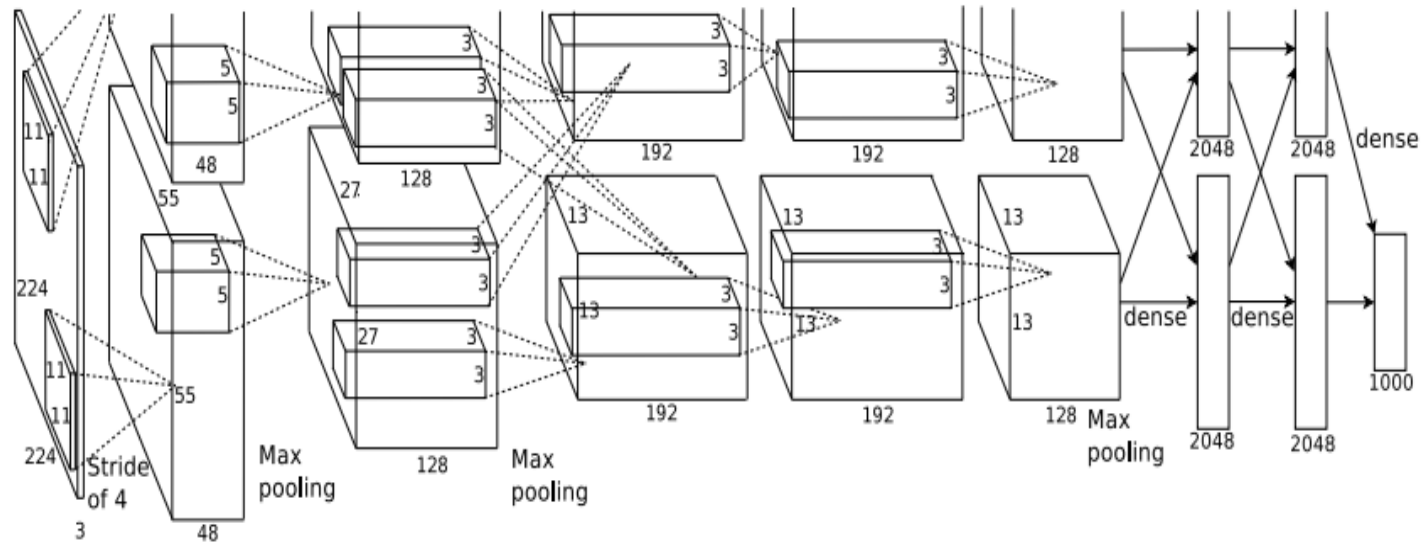


San Marco Square: 13,699 images, 4,515,157 points

Agarwal et al. Building Rome in day. ICCV, 2009.

Histoire de la vision par ordinateur

- 2010s
 - Deep Learning in CV

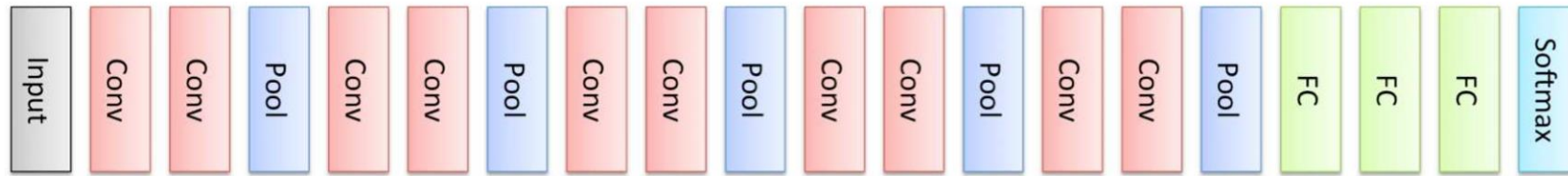


AlexNet. Krizhevsky et al., 2012, designed for ImageNet classification

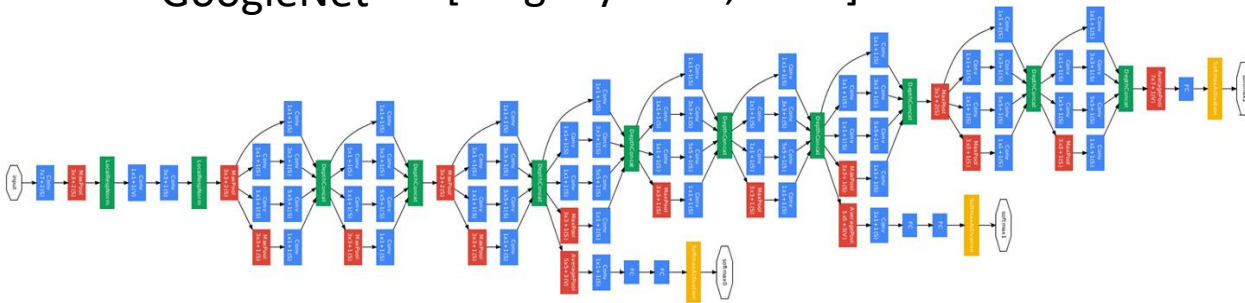
Histoire de la vision par ordinateur

- 2010s
 - Réseaux plus profonds et plus larges,

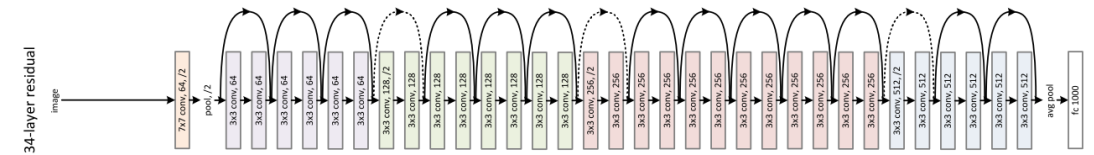
VGGNet [Simonyan and Zisserman, 2014]



GoogleNet [Szegedy et al., 2014]

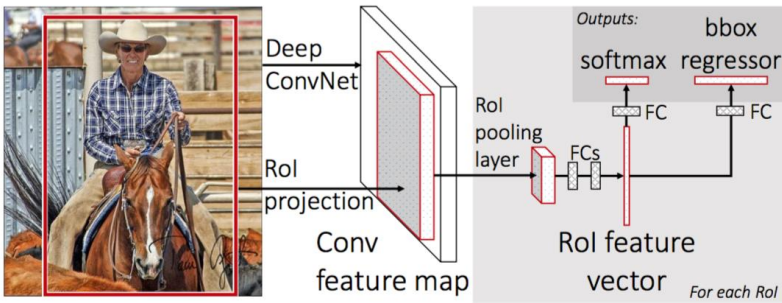


ResNet [He et al., 2015]

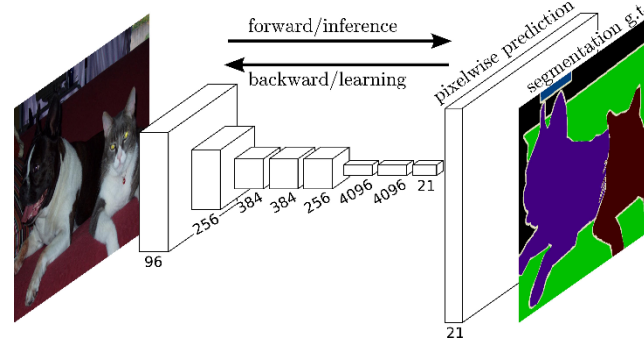


Histoire de la vision par ordinateur

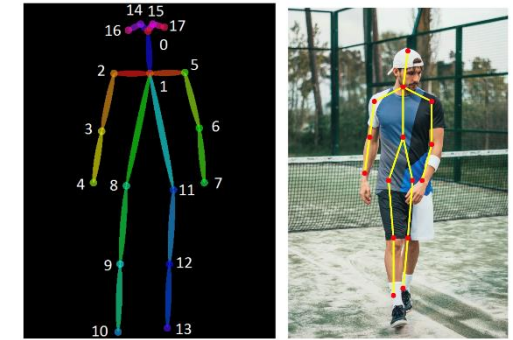
- 2010s
 - Réseaux neuronaux pour la reconnaissance.



Object Detection (Fast RCNN, Girshick, 2015)



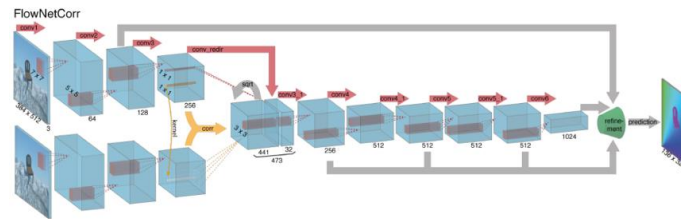
Semantic Segmentation (FCN, Long et al., 2014)



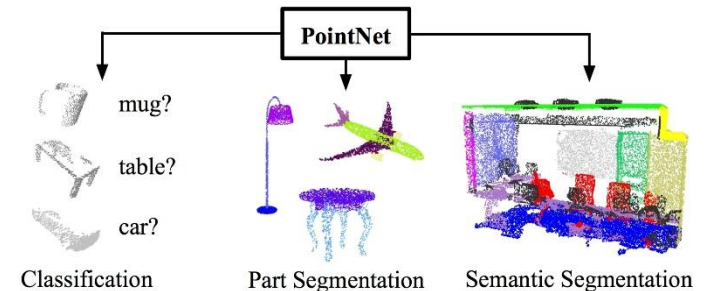
Human Pose Estimation (OpenPose, Cao et al., 2017)



Depth Estimation (Eigen et al. 2014)



Optical Flow (FlowNet Fischer et al. 2015)



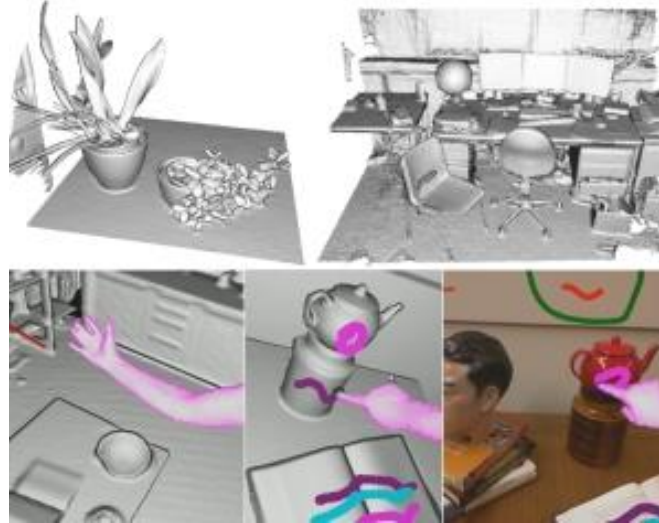
Point Cloud Recognition (PointNet, Qi et al., 2016)

Histoire de la vision par ordinateur

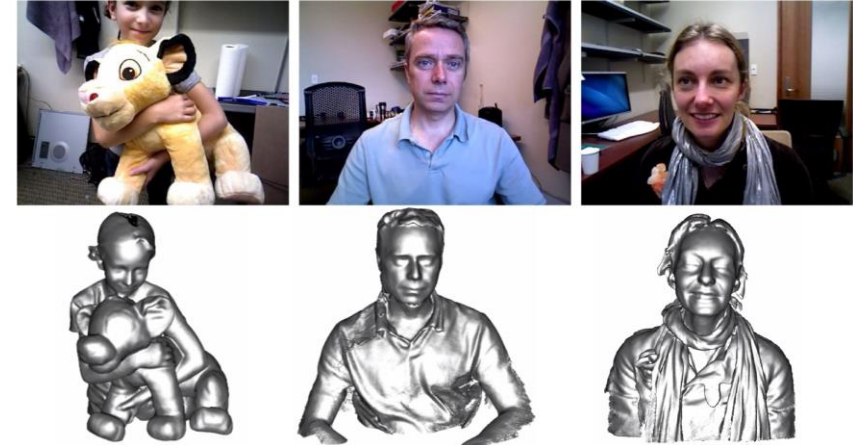
- 2010s
 - Détection de profondeur et vision 3D.



Microsoft Kinect, 2010



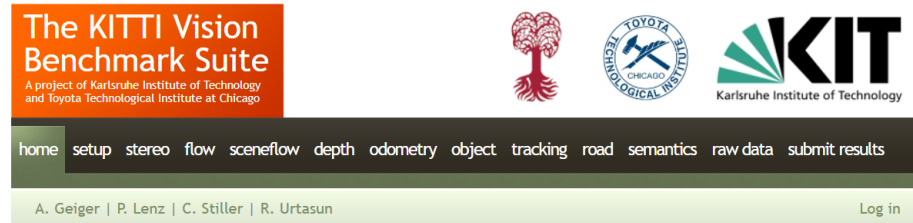
KinectFusion, Newcombe et al., 2011



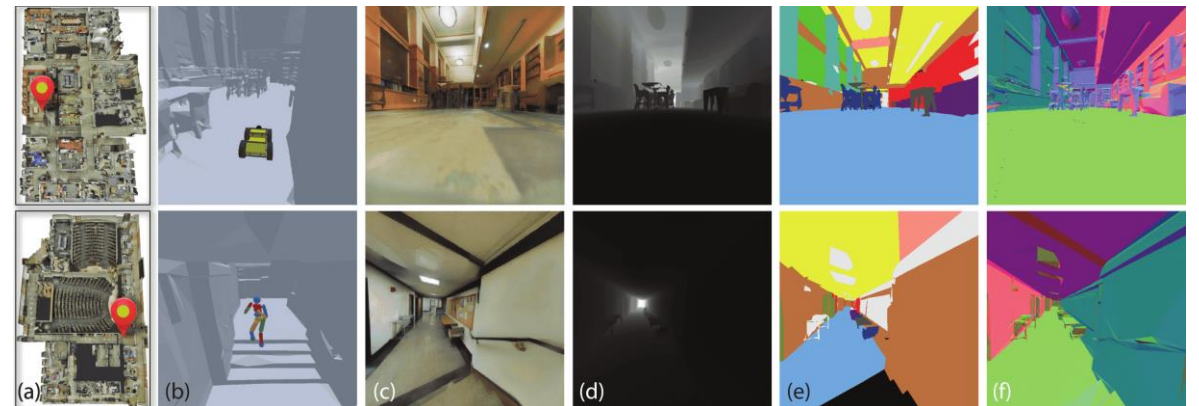
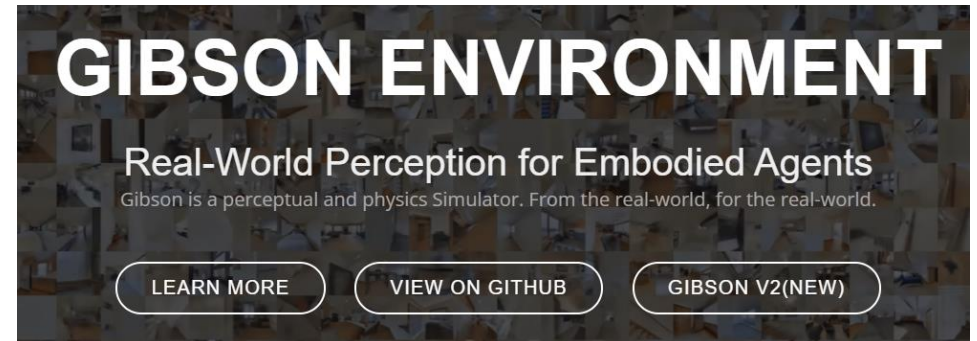
DynamicFusion, Newcombe et al., 2015

Histoire de la vision par ordinateur

- 2010s
 - Conduite autonome et Systèmes Incarnés (Embodied Systems) :



The KITTI dataset, Geiger et al., 2012

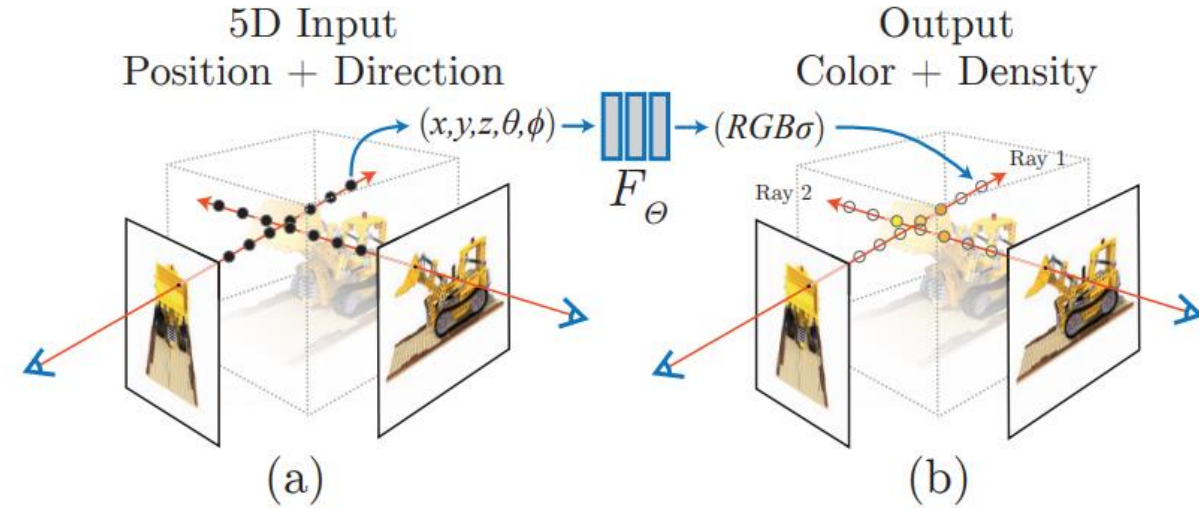
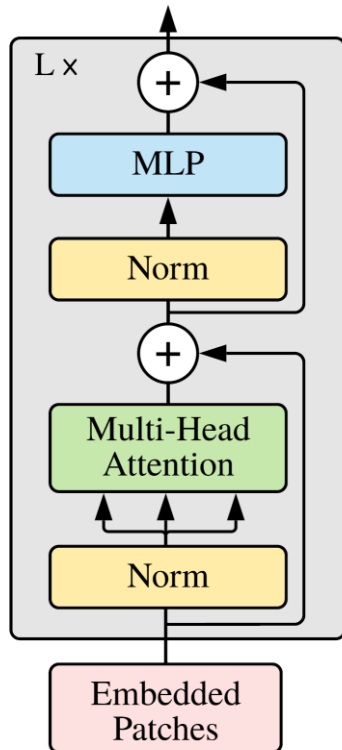


The Gibson environment, Xia et al., 2018

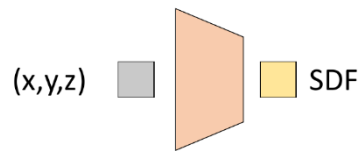
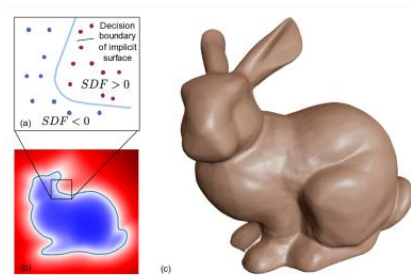
Histoire de la vision par ordinateur

- 2010s
 - Neural implicit representations
 - Vision transformers,
 - Mécanisme d'Attention
 - Fine-Tuning

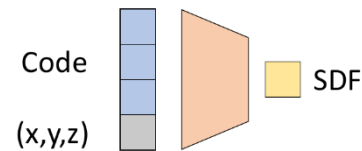
Transformer Encoder



NeRF: Neural Radiance Fields. Midenhall et al. 2020



(a) Single Shape DeepSDF



(b) Coded Shape DeepSDF

DeepSDF, Park et al., 2019

Histoire de la vision par ordinateur

- 2023
 - OpenAI GPT (ChatGPT)
 - Sora AI Video Generator
 - Large Multi-modal Model
 - Etc,

give me an image showing a computer vision field



Here is an image representing the computer vision field, illustrating key elements such as 3D model reconstruction, neural networks, and multi-view camera setups.

Description d'un système de Vision

On trouve dans les systèmes de vision deux catégories distinctes :

Systemes bidimensionnels: Ce sont les systèmes qui n'incluent pas la notion de **profondeur** ; c'est à dire ils ne peuvent reconnaître que des objets plans puisqu'ils n'utilisent que des données à deux dimensions. Ces systèmes trouvent place dans la télédétection, la biomédicale, la reconnaissance des caractères, etc

Systemes tridimensionnels: Ces systèmes sont pour leur part embarqués sur les robots mobiles évoluant en scènes d'intérieur ou d'extérieur, les dotant ainsi d'un sens qui contribue à une autonomie dans la prise de décision imitant pour cela le système visuel humain comme par exemple : Guidage d'un bras d'un robot, l'évitement d'obstacles.

Les systèmes actifs

Ils prennent en charge le contrôle de l'éclairage de la scène. Ils utilisent des capteurs de type laser (télémètre laser) ou ultrason .

Ils disposent d'un émetteur et d'un récepteur qui acquière une information spatiale

La triangulation : le principe se fonde sur la connaissance de la distance entre le récepteur et l'émetteur, ainsi on peut déduire la profondeur du point qui réfracte l'onde de la direction du signal émis et réfracté.

Le temps de vol : on peut calculer la profondeur du point par la mesure du temps mis par le signal entre l'émetteur et le récepteur.

Les systèmes passifs :

Ils ne contrôlent pas l'éclairage de la scène.

Les seules données disponibles sont les images (donc le degré de complexité est élevé par rapport aux systèmes actifs).

Ils traitent une ou plusieurs images 2D prises par des caméras et contenant une information de luminance (niveau de gris ou couleur) émise ou réfléchi par l'environnement

Modèle de David Marr

Modèle de MARR (MIT 1982)

- Le Modèle de David Marr en vision par ordinateur est une approche théorique qui vise à expliquer comment les systèmes de vision, qu'ils soient biologiques ou artificiels, perçoivent et interprètent les informations visuelles.
- Marr a formulé ce qu'il a proposé une analyse en trois niveaux, chacun se concentrant sur différents aspects du traitement visuel.
- Le modèle de Marr a profondément influencé la recherche en vision par ordinateur, en fournissant une structure pour développer des algorithmes de traitement d'images qui imitent le fonctionnement du système visuel humain. Des techniques comme la détection de contours, la segmentation d'images et la reconstruction 3D tirent parti des concepts formulés par Marr.

Modèle de David Marr

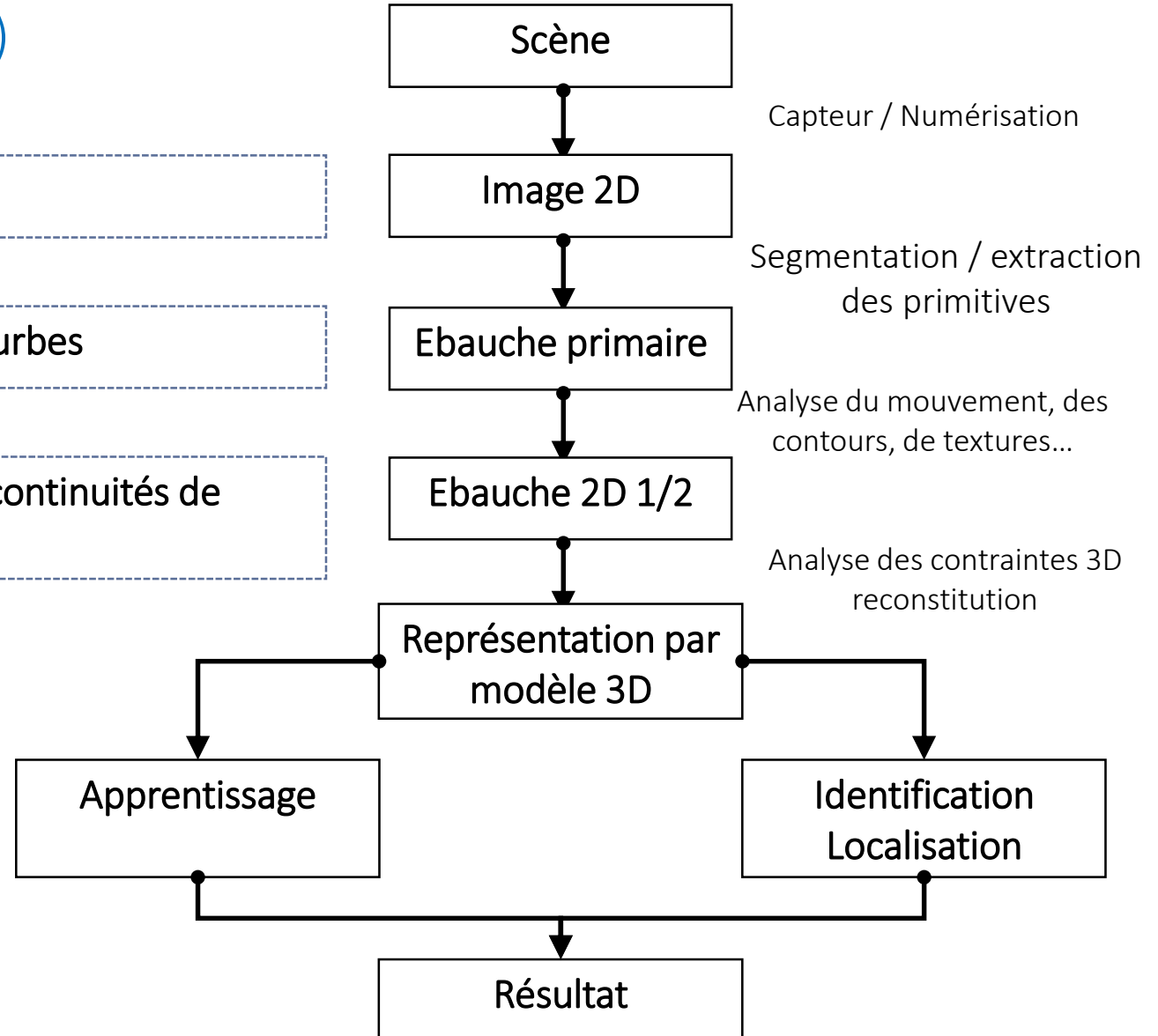
Un ensemble d'ébauches (croquis)

Intensités perçues

Contours, Lignes, Points, Courbes

Orientation des surfaces locales, discontinuités de profondeur

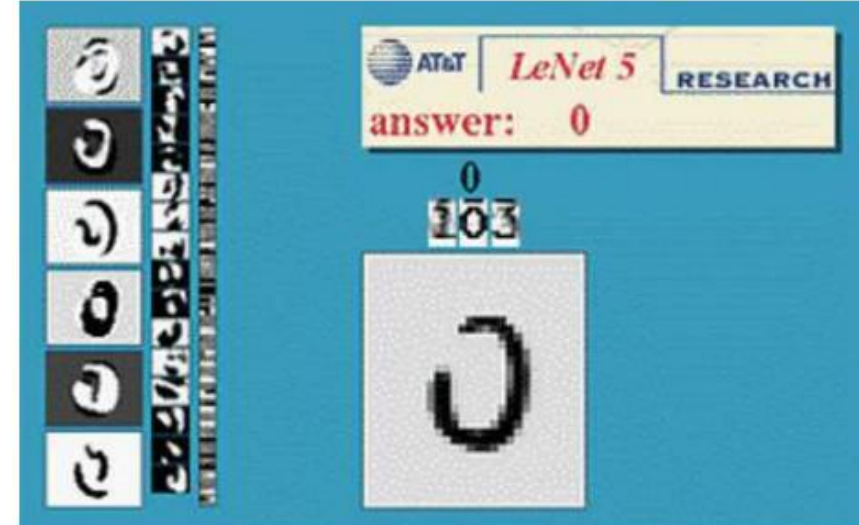
Modèle 3D organisé en primitives de surfaces et volumes



Domaines d'application

La vision par ordinateur est utilisée aujourd'hui dans une grande variété d'applications de la vie quotidienne, notamment :

- **Optical character recognition (OCR):** Lecture des codes postaux manuscrits sur les lettres et reconnaissance automatique des plaques d'immatriculation.
- **Machine inspection:**
 - Le scanne rapide des pièces pour l'assurance qualité à l'aide de la stéréo vision avec un éclairage spécialisé pour mesurer les tolérances sur les ailes d'avion ou les pièces de carrosserie automobile
 - détection des défauts dans les pièces en acier moulé à l'aide de la vision par rayons X.
- **Commerce :** reconnaissance d'objets pour les caisses automatiques et les magasins entièrement automatisés.

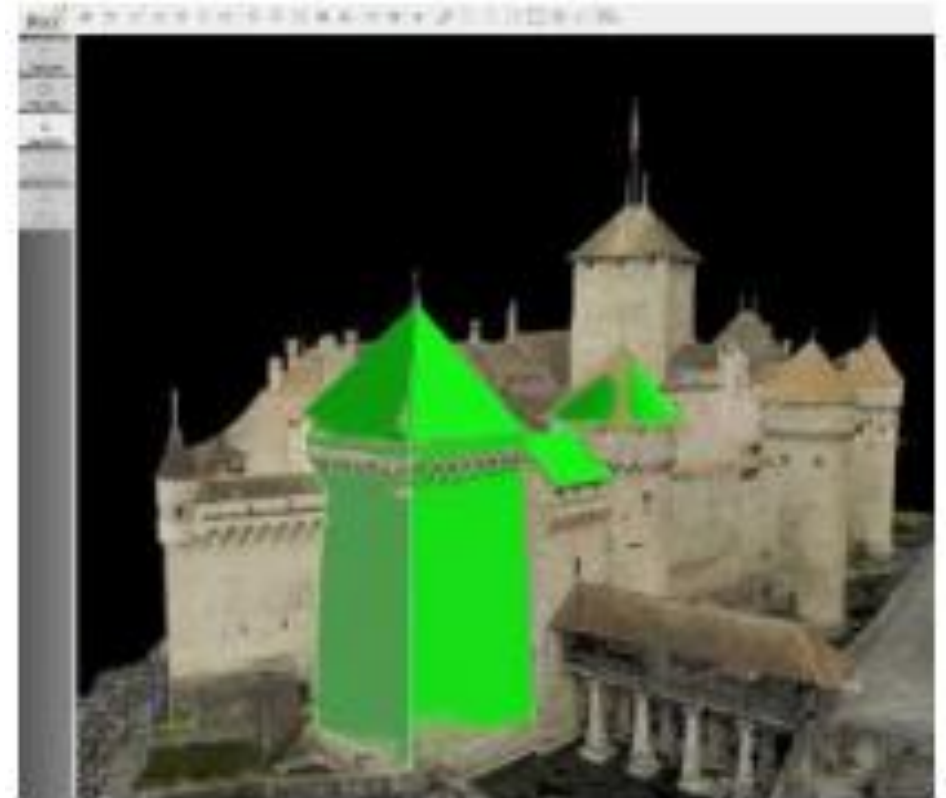


Domaines d'application

- **Logistique d'entrepôt** : prélèvement de pièces par des manipulateurs robotiques.
- **Imagerie médicale** : enregistrer les images avant et pendant l'opération ou étudier la morphologie du cerveau des personnes au fil du temps.
- **Véhicules autonomes** : capables de conduire d'un point à un autre entre les villes.



- **Construction de modèles 3D (photogrammétrie) :** construction entièrement automatisée de modèles 3D à partir de photographies aériennes et de photos de drones.
- **Match move:** fusionner des images générées par ordinateur (CGI) avec des séquences vidéo en direct en suivant des points caractéristiques dans la vidéo source pour estimer le mouvement de la caméra en 3D et la forme de l'environnement.



Domaines d'application

- **Capture de mouvement:** utilisation de marqueurs rétro-réfléchissants vus depuis plusieurs caméras ou d'autres techniques basées sur la vision pour capturer les acteurs pour l'animation par ordinateur ;
- **Surveillance :** surveillance des intrus, analyse du trafic autoroutier.
- **Reconnaissance d'empreintes digitales et biométrie :** pour l'authentification automatique d'accès ainsi que pour des applications judiciaires.
- **Vidéos 3D (360°) :** (exemple YouTube)
<https://www.youtube.com/watch?v=sPyAQQk1c1s&list=PLe2QTEWuVb-uYqfd7iBfnjOFZu-cxl6Cw>

La réalité virtuelle

La réalité virtuelle (virtual reality) est la technologie qui permet de plonger/mettre une personne dans un monde artificiel (numérique).

exemple YouTube : <https://www.youtube.com/watch?v=ArlavrOFpkl>



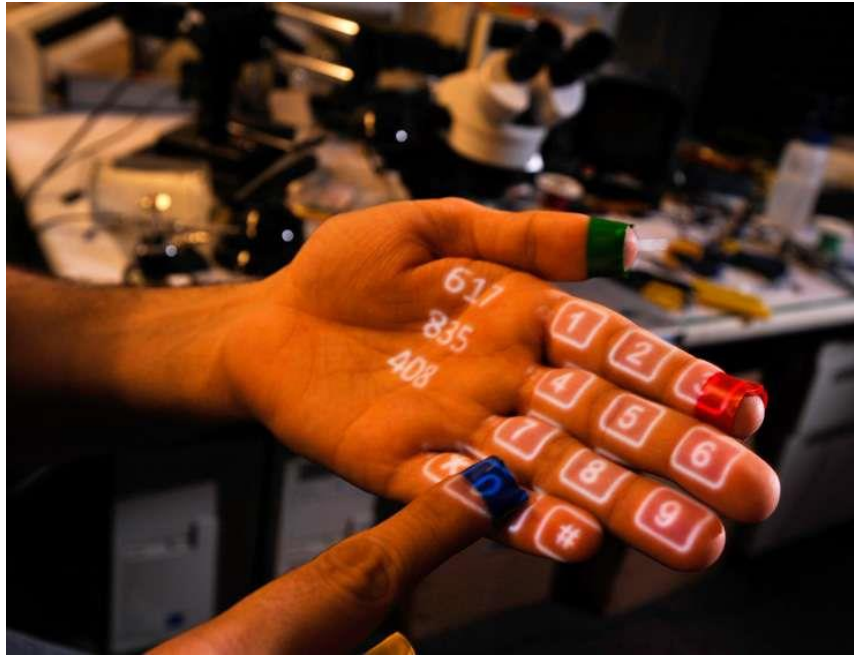
Src : <https://www.artefacto-ar.com/realite-virtuelle/>



Source : Dieter Schmalstieg & Axel Pinz, *Augmented Reality*, ICG (Allemagne).

La réalité augmentée

La réalité augmentée permet la superposition d'éléments virtuels sur des éléments réels.

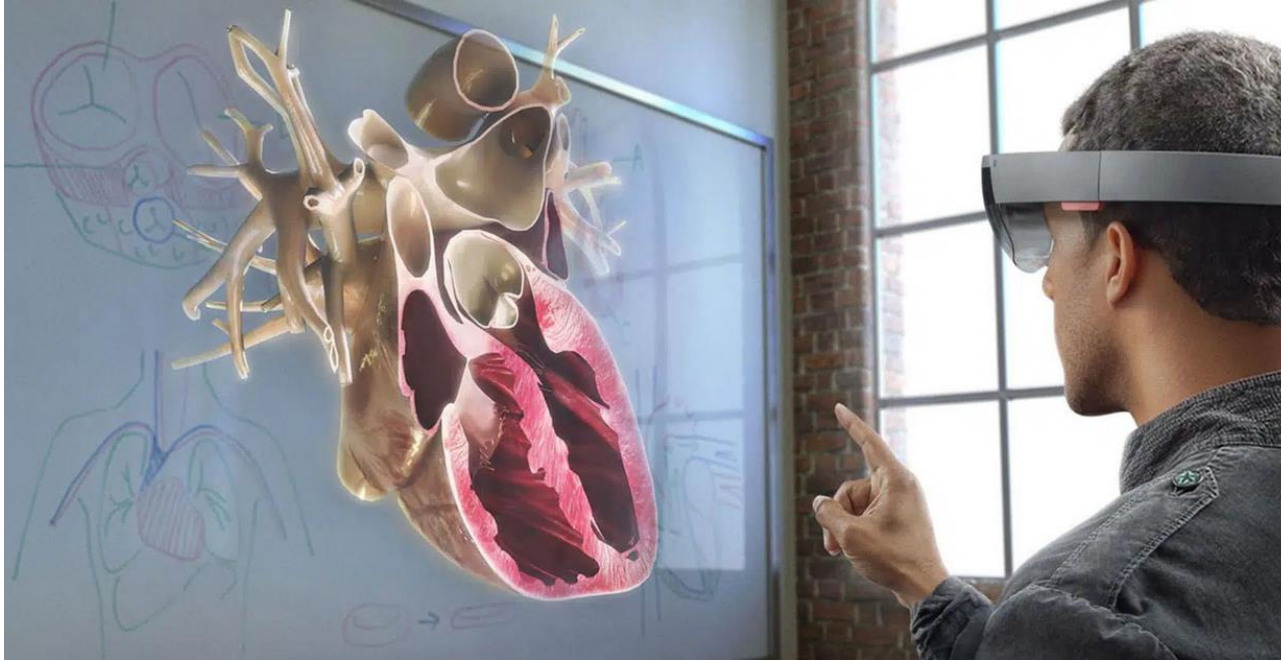


Src : <https://www.futura-sciences.com/tech/dossiers/technologie-realite-augmentee-applications-977/>



Src : <https://www.anthedesign.fr/autour-du-web/realite-augmentee-ra/>

La réalité augmentée

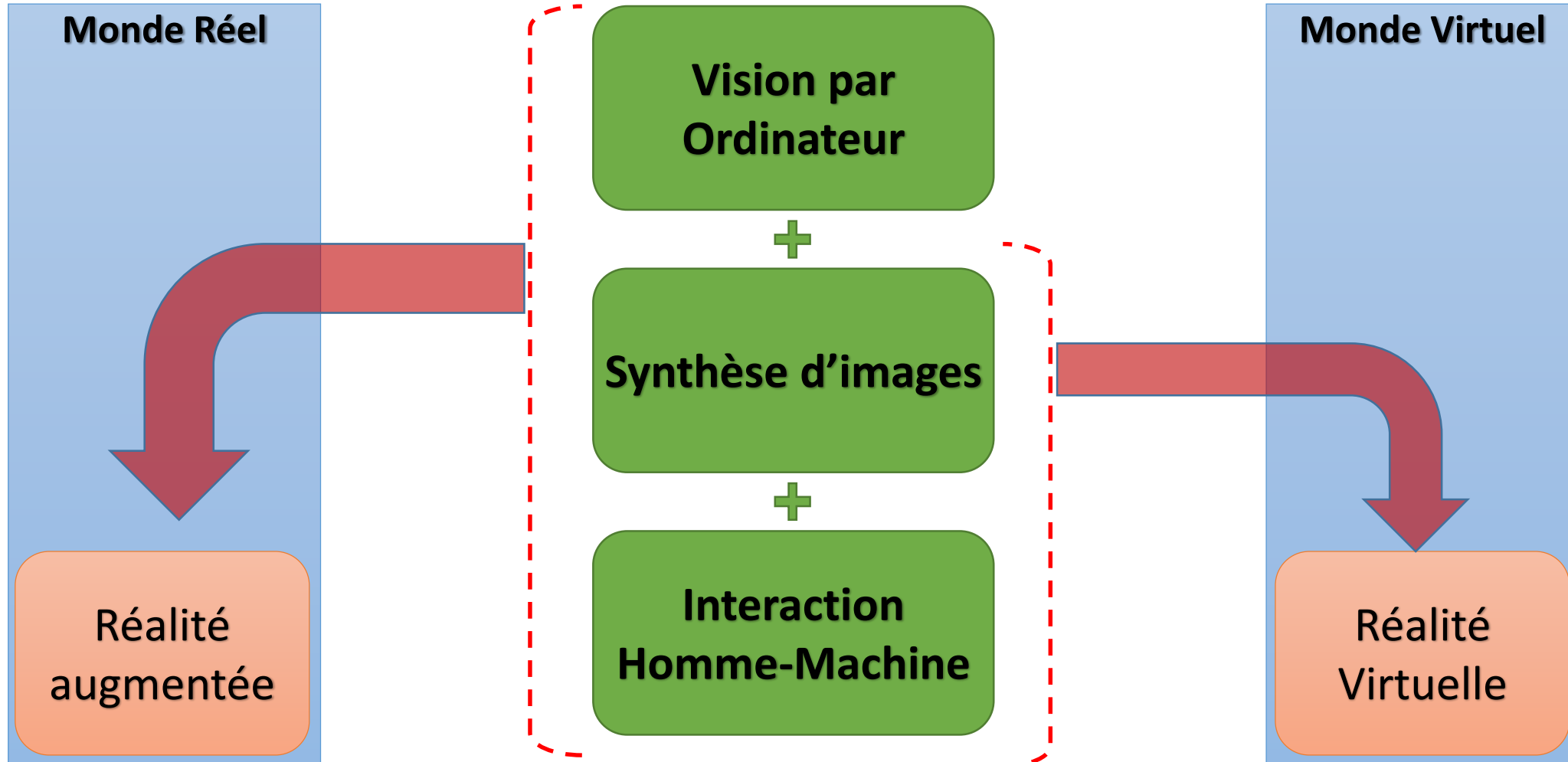


Src : <https://www.anthedesign.fr/autour-du-web/realite-augmentee-ra/>



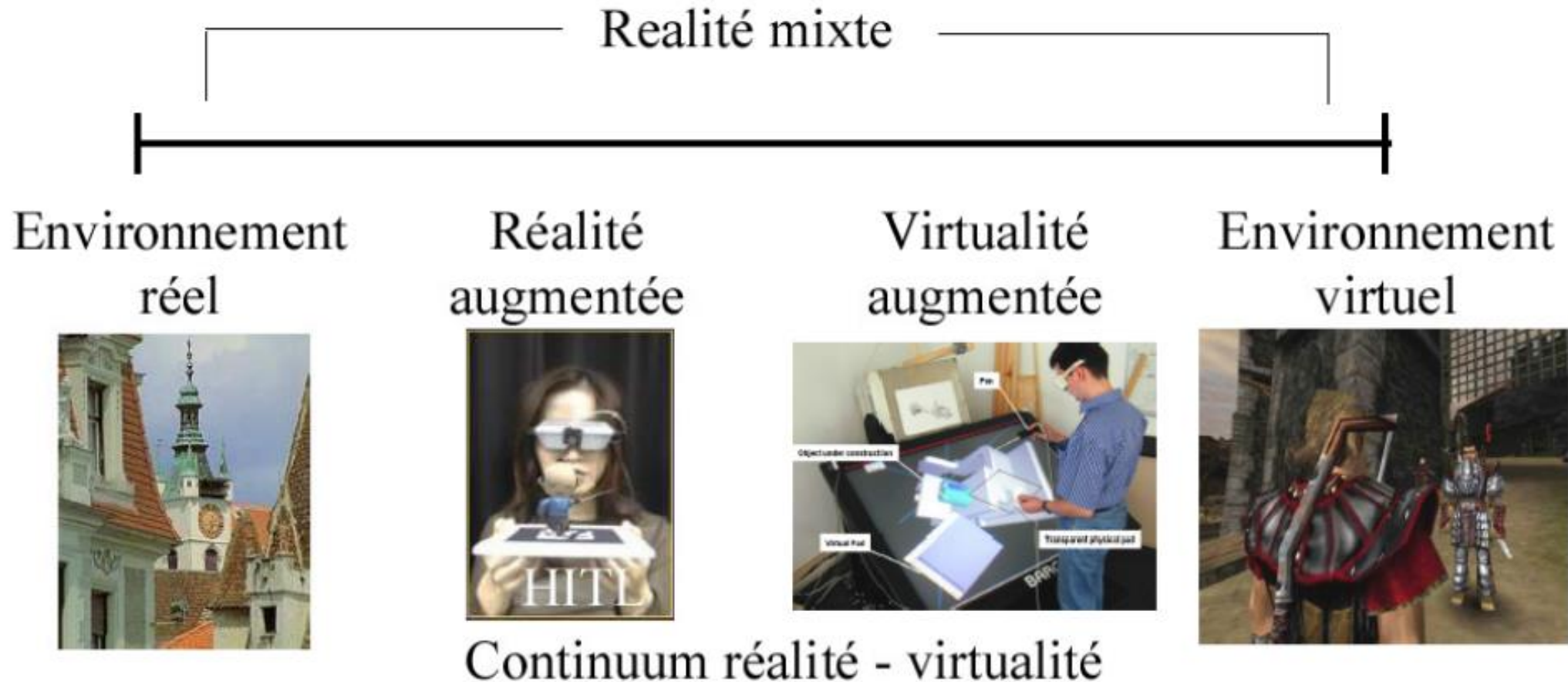
<http://blog.ametracgroup.com/maintenance-facilitee-par-la-realite-augmentee/>

La réalité virtuelle/ augmentée



La réalité virtuelle/ augmentée

On parle parfois de **réalité mixte** (*mixed reality*) pour inclure toutes les variantes du réel au virtuel



La réalité virtuelle/ augmentée



Visiocasque



Gant

Combinaison



Vision cognitive

Nouveau sous-domaine de la vision par ordinateur :

Vision par ordinateur + Sciences cognitives = Vision Cognitive

Sciences cognitives : Etude du cerveau humain, de l'intelligence, de la connaissance

Vision cognitive (par ordinateur) : intégration des processus intelligents dans la vision par ordinateur

L'interprétation d'une image est difficile Nous utilisons beaucoup d'informations a priori pour reconnaître une scène.

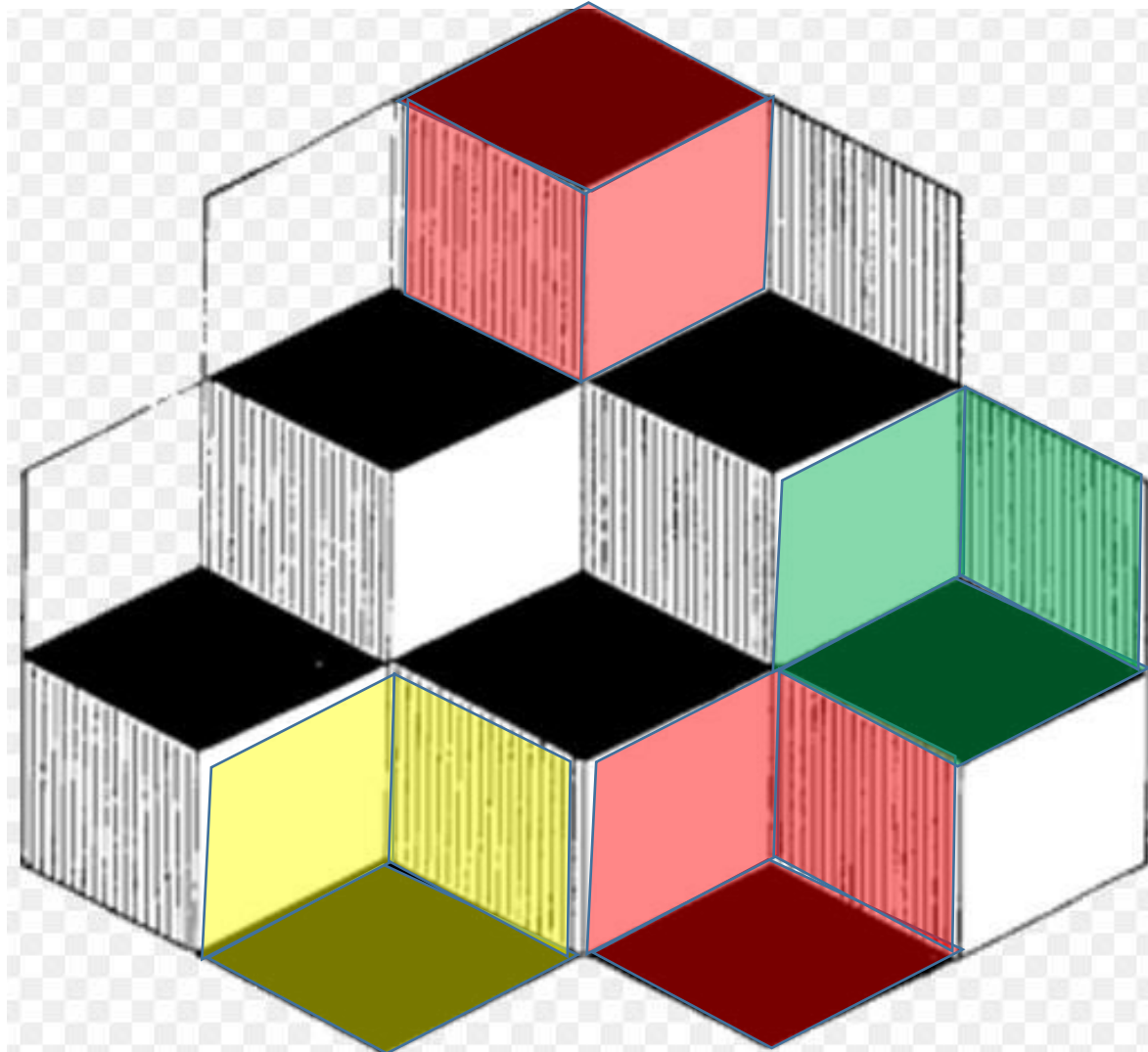
Comment le cerveau humain interprète t-il les images?

Vision cognitive

- Le cerveau humain interprète bien le contenu des images mais il n'est pas assez précis pour mesurer les grandeurs et la taille des objets (très bonne reconnaissance).
- Le cerveau humain arrive à voir plus que le contenu brut des images
- Le cerveau humain est massivement parallèle, ce qui compense sa lenteur
- La machine interprète mal (scène 3D en 2D, Les informations des pixels ne sont pas suffisantes, etc...)
- La machine interprète moins bien, mais effectue des mesures plus précises.

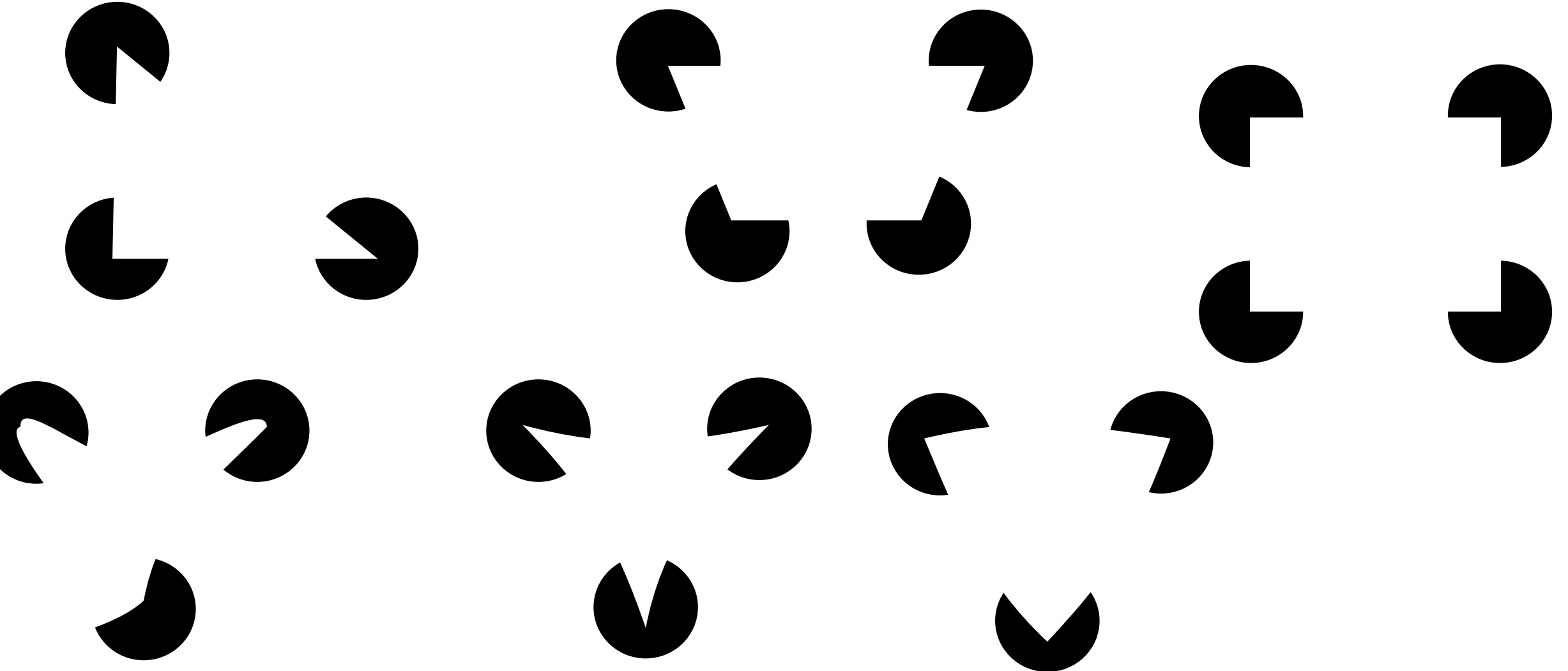
Exemple: Une maison, une chaise, etc... pour l'homme est une notion abstraite, mais interprétation différente pour la machine (RVB= XXX, Texture = YYYY, couleur = ZZZZ, etc...)

Combien y a-t-il de cubes dans cette image?

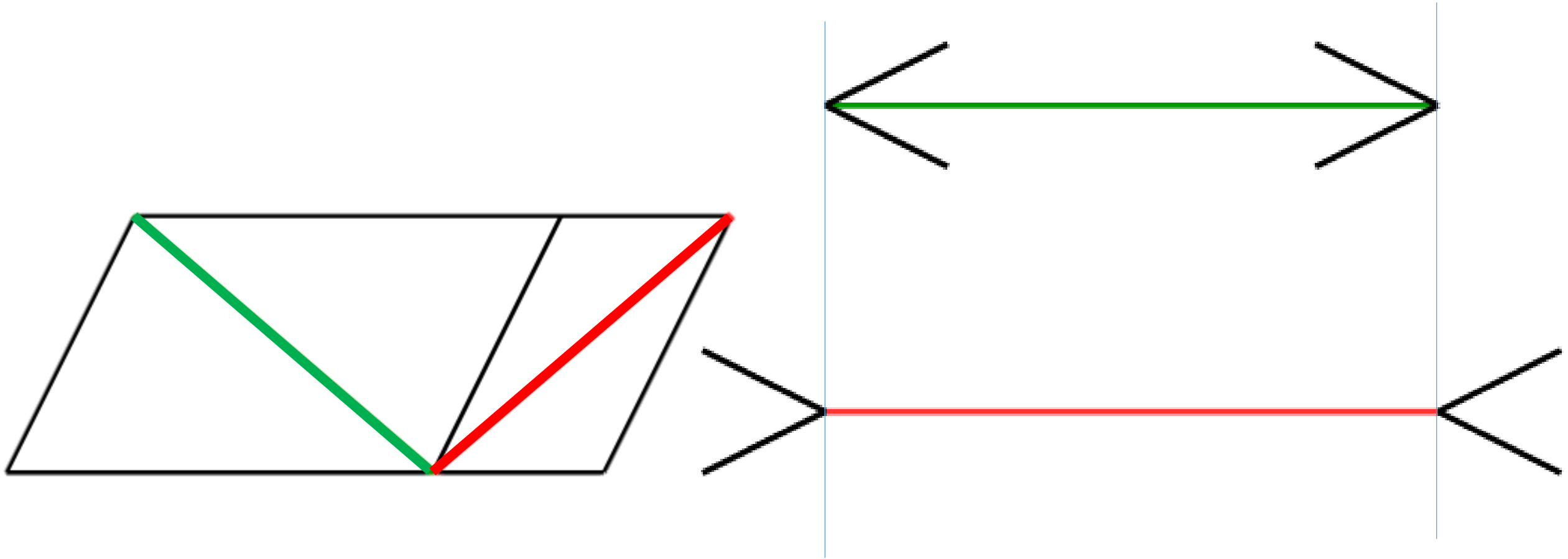


La réalité virtuelle/ augmentée

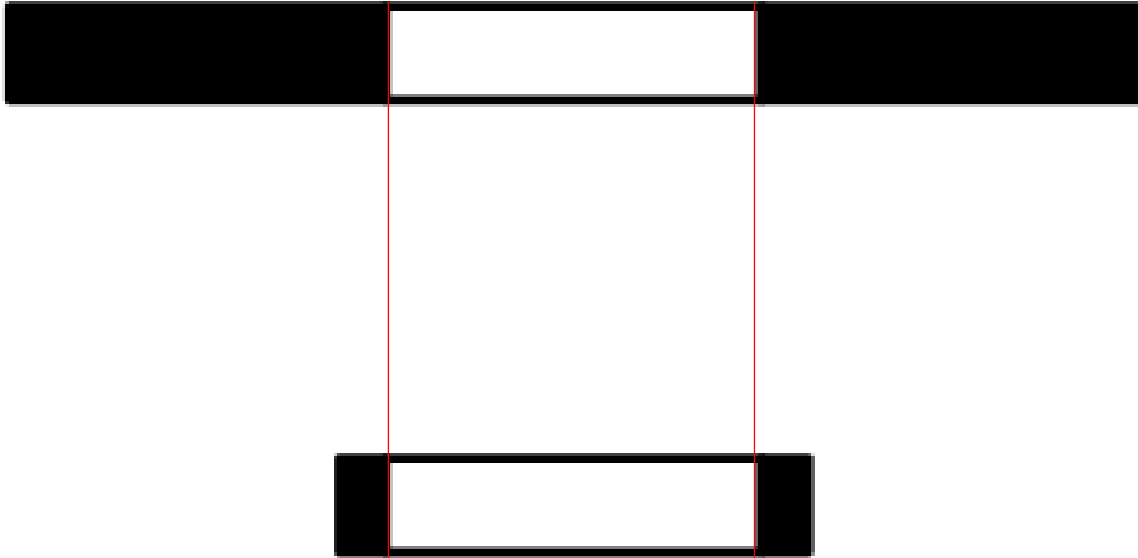
Le cerveau interprète ce qu'il voit en fonction de ce qu'il connaît.



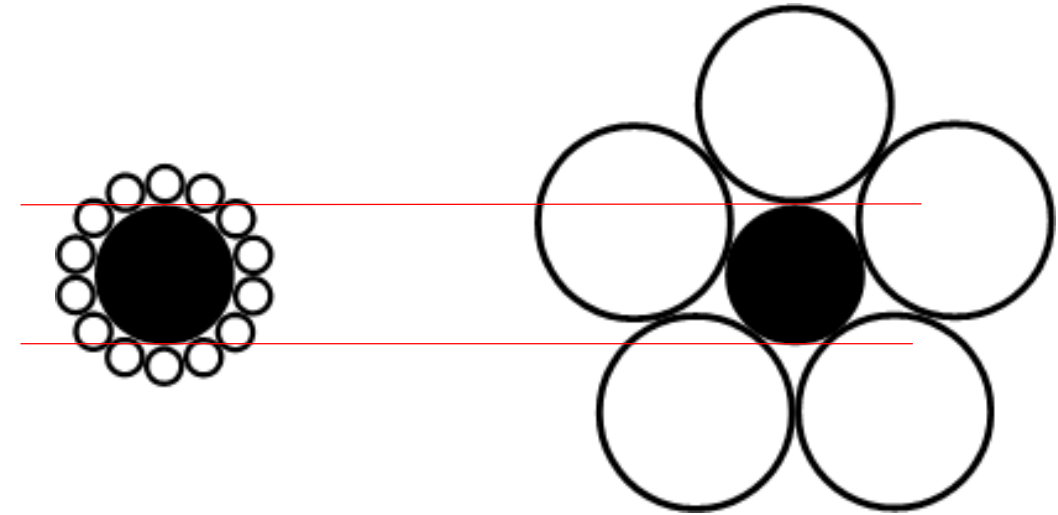
Illusions de dimensions !!!



Illusions de dimensions !!!

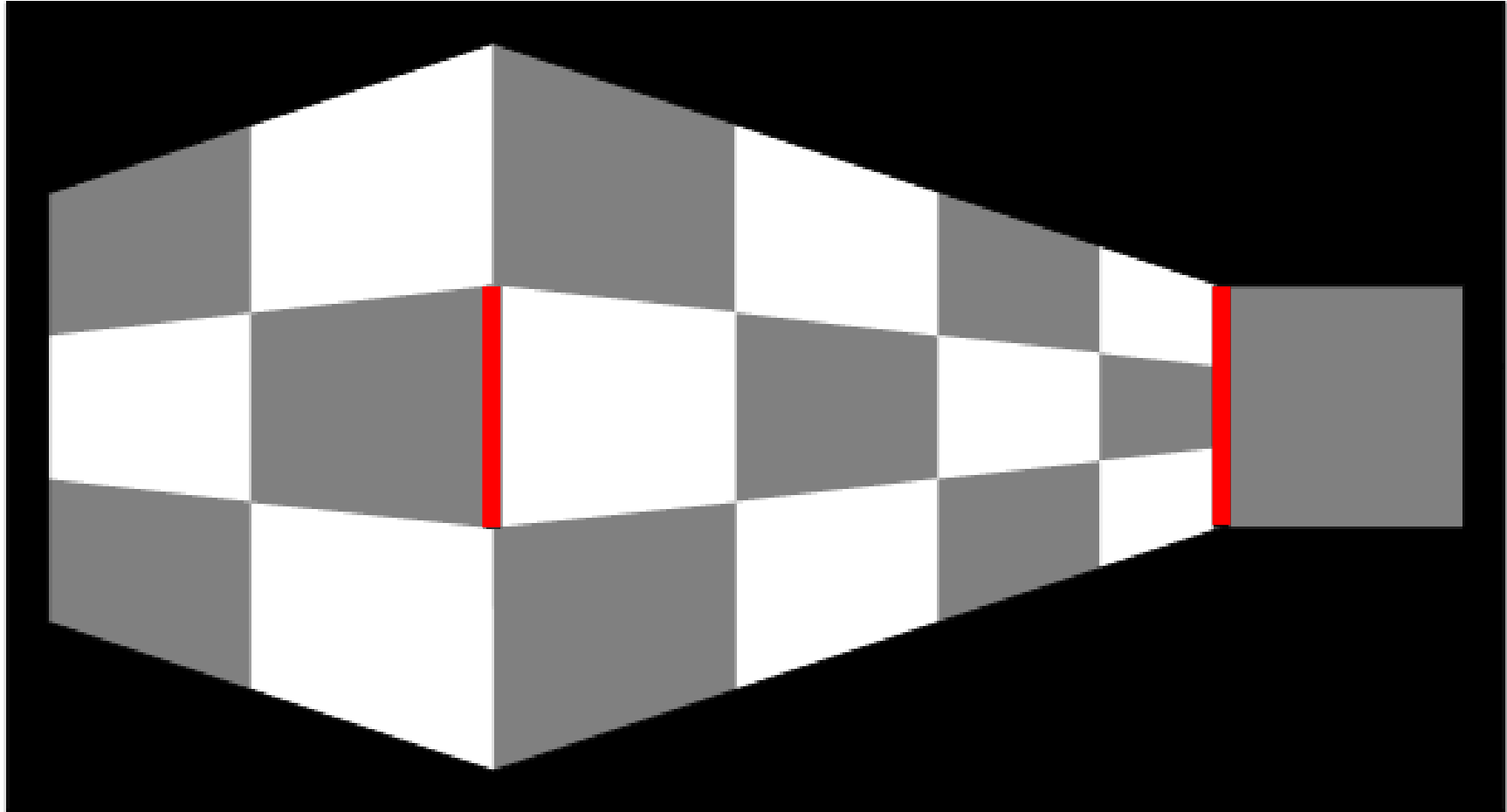


les rectangles blancs sont exactement de la même taille



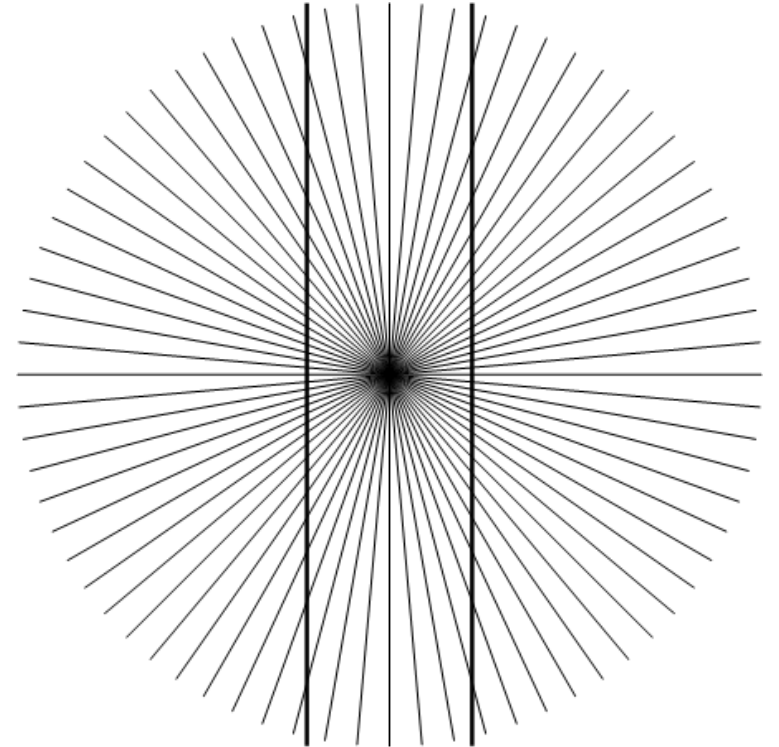
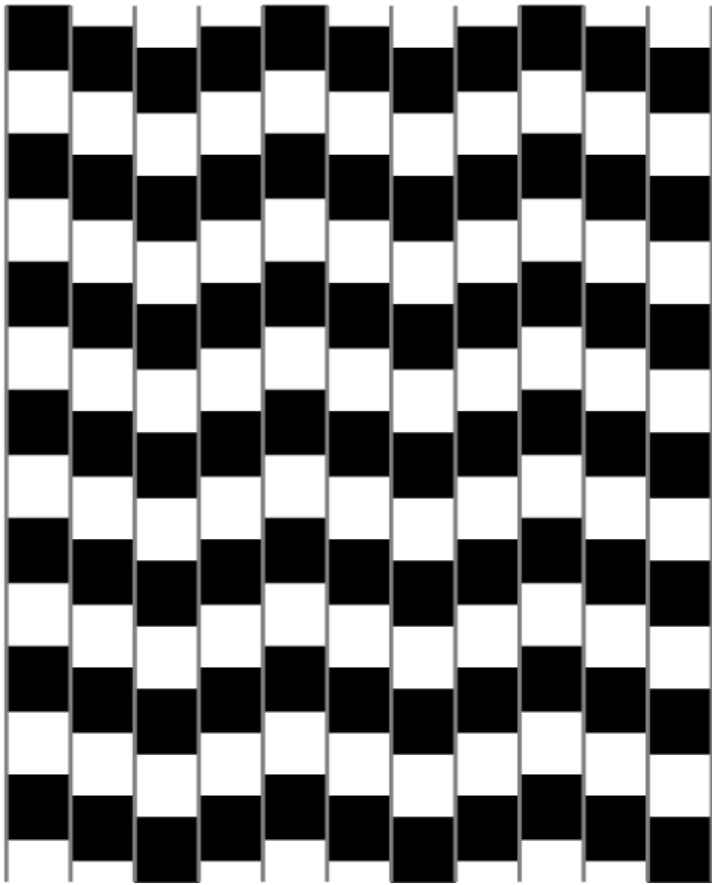
les disques noirs sont exactement de la même taille

Illusions de dimensions !!!

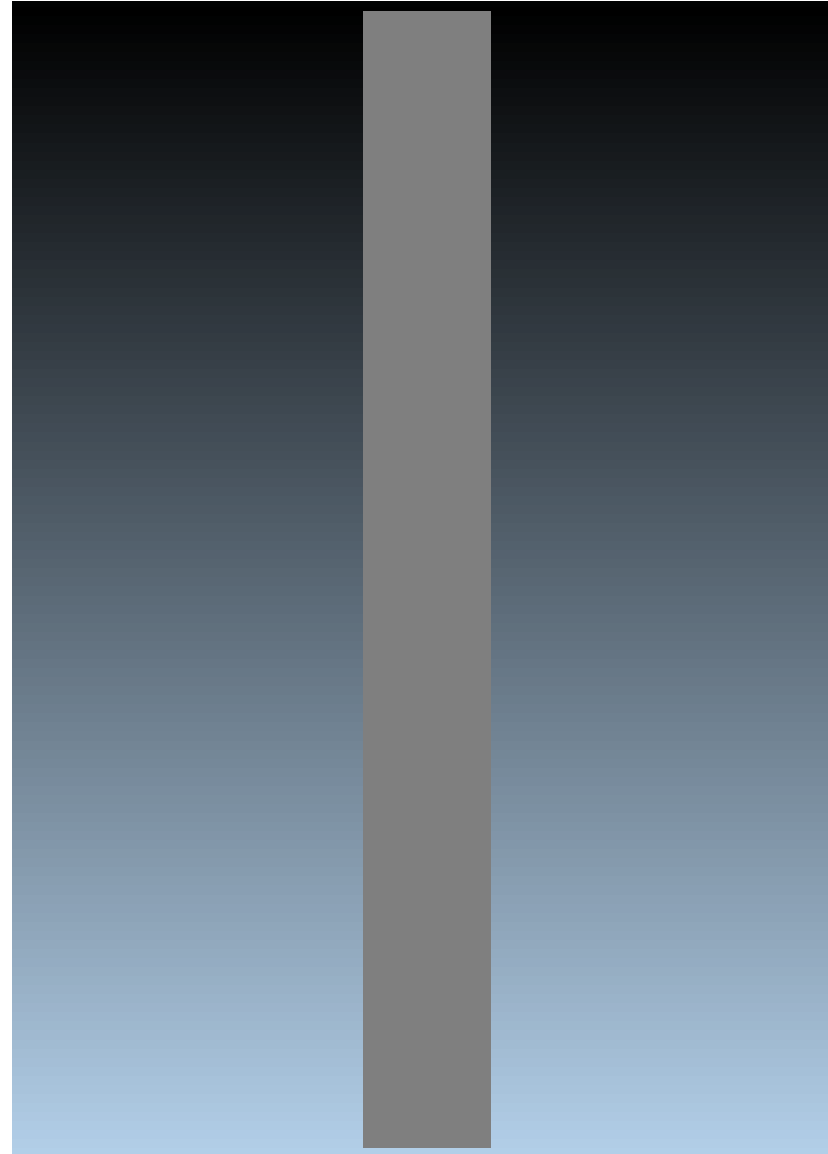


Les segments sont exactement de la même taille

Illusions de directions



Illusions de couleur



La couleur de segment est parfaitement uniforme

Illusions de dimensions !!!

