

Computer Vision

Raphaël Viards – Magellium – Imagerie et Applications

raphael.viards@magellium.fr

Computer Vision

- Introduction
- Perception 3D
- Analyse de l'environnement
- Détection et Localisation

Introduction

- I. Magellium
- II. Métier et Projets
- III. Computer Vision



Observation
de la terre



Géo-information



Imagerie
& Applications

A circular image containing a snippet of software code written in a programming language like JavaScript or Java.

Applications
logicielles



Infrastructure
logicielle

1998
Création

2 sites
Paris
Toulouse

248
employés

26 M€
CA 2019



Observation
de la terre



Géo-information



Imagerie & Applications

2003
Création

170+
employés

2 sites
Paris
Toulouse

17 M€
CA 2018

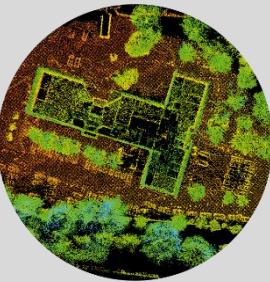
Intégrée au groupe ARTAL depuis le
1^{er} Septembre 2016



Image & vidéo



Lidar & 3D



Robotique



Activités

29

Docteurs &
Ingénieurs

Offre



Transfert de
technologies

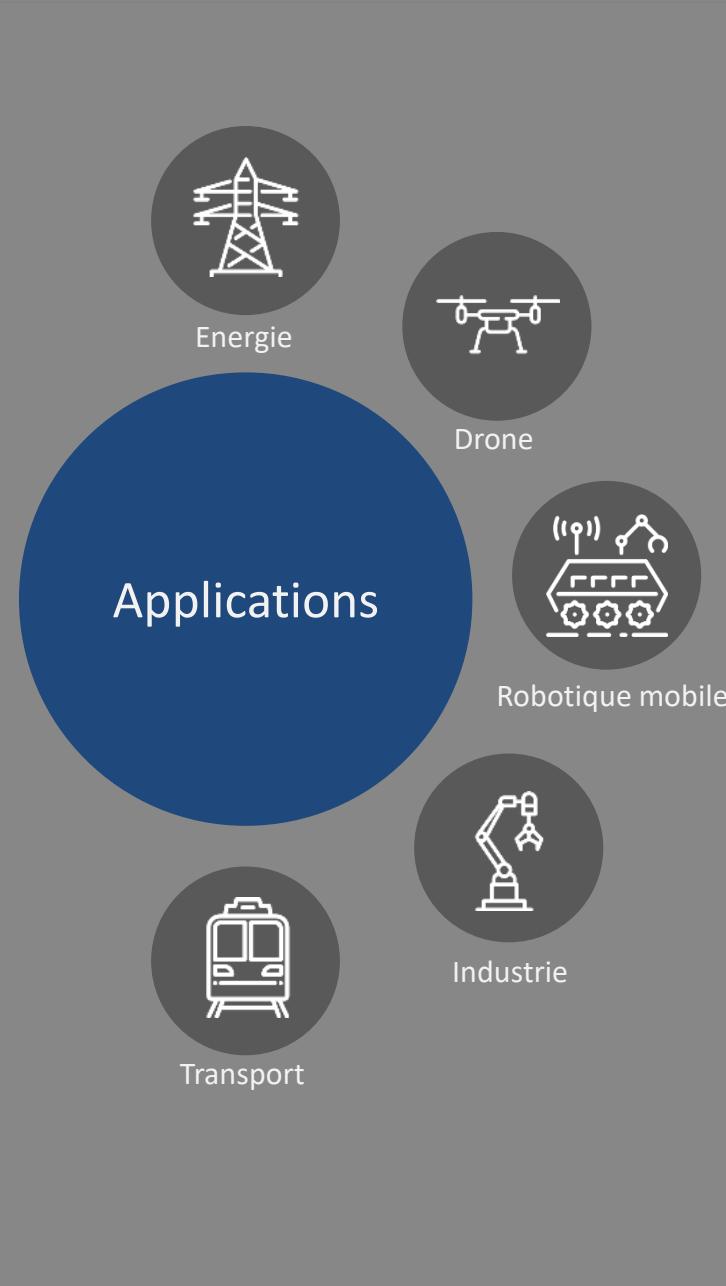
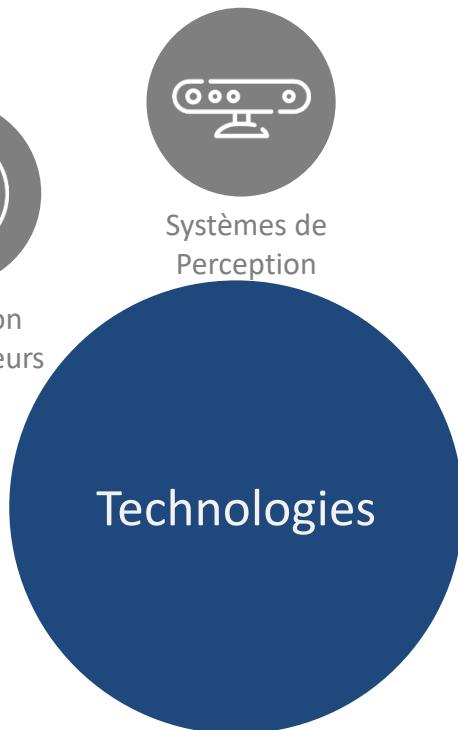


Conception
systèmes de perception



Développement
logiciel

3,1 M€
CA 2019





Capteurs 2D & 3D



Détection et Reconnaissance d'objets



Localisation et navigation



Technologies informatiques avancées

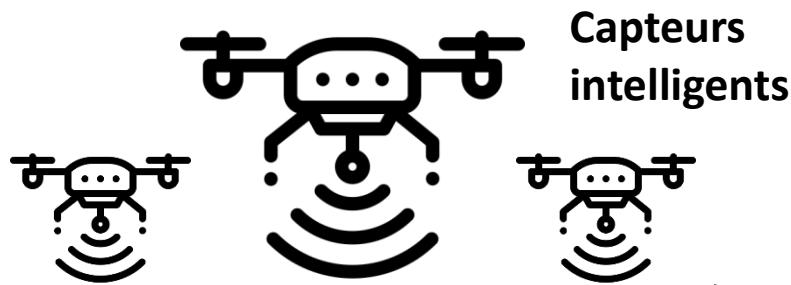
Cameras matricielles (optique, IR)
Cameras actives (RGB-D, Time-of-flight)
Stéréovision, LiDAR, profil laser

Reconnaissance de formes 2D/3D
Détection du changement, classification
Traitement nuages de points denses
Reconstruction & visualisation 3D

Odométrie visuelle, SLAM, flot optique
Hybridation IMU / Vision / GPS
Intégration vecteur (eau, air, terre)

Programmation parallèle (GPU)
Calcul embarqué (Linux, ARM)
Deep Learning (Tensor Flow)
Technologies BigData (HDFS, cloud)

Synthèse



Station sol ergonomique



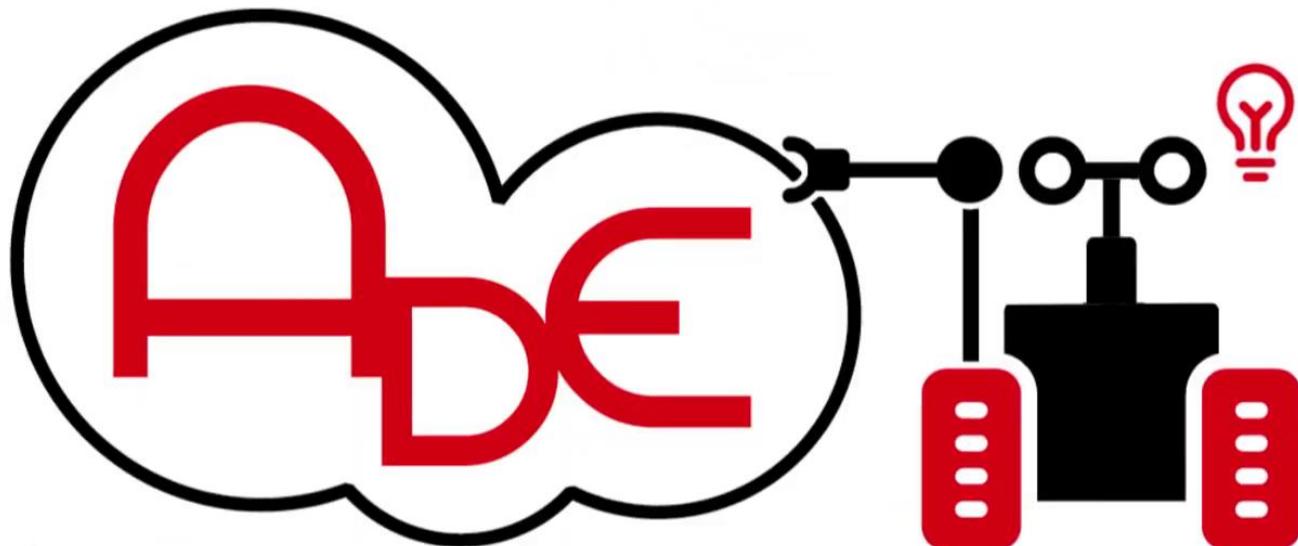
Introduction

- I. Magellium
- II. Métier et Projets
- III. Computer Vision

- Responsable technique
 - Définition des architectures et algorithmes
 - Encadrement technique des équipes
 - Échanges clients et/ou partenaires
- Thématiques:
 - Computer vision
 - Robotique d'exploration planétaire
 - Robotique orbitale
- Autres:
 - Labo vision
 - Deep learning



Data Fusion for Space Robotics





IN-SPACE ASSEMBLY SIMULATION

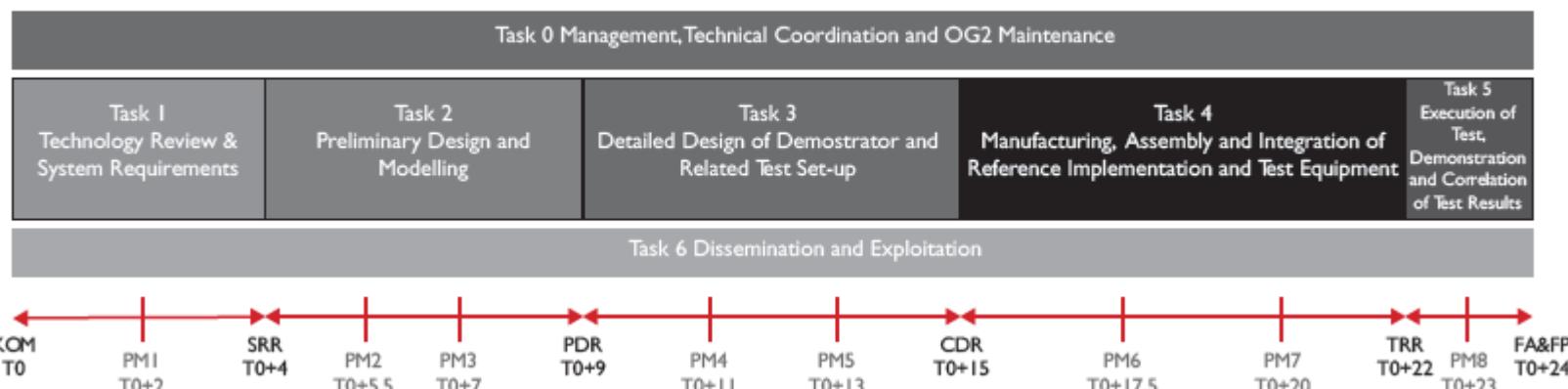
OBJECTIVE : Development, evaluation, and demonstration in space conditions

This demonstrator aims to simulate a fully representative scenario with spacecraft AOCS considerations, and serves as a development and prototyping tool to validate interfaces between software components

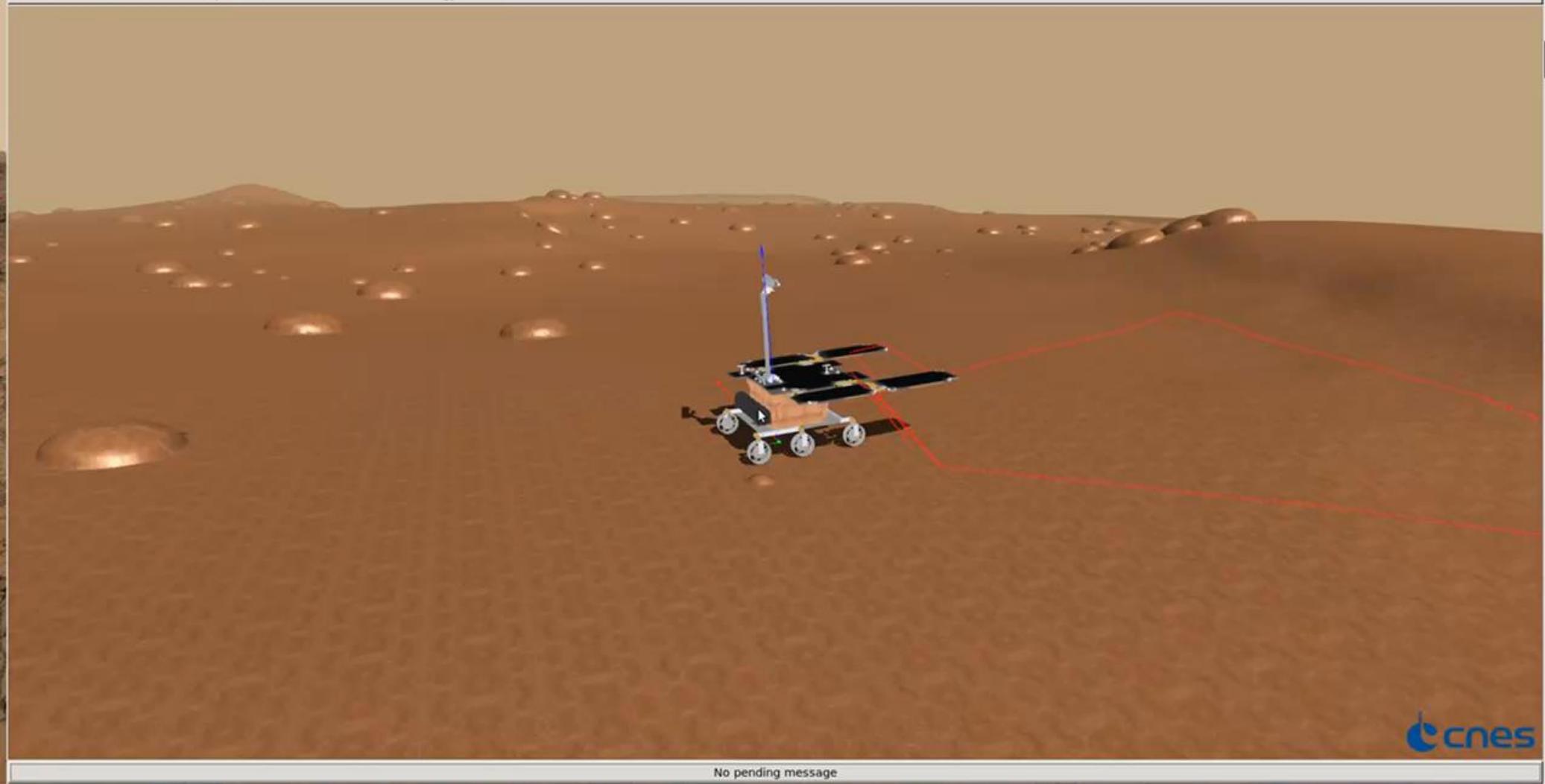
CoRob-X

Cooperative Robots for Extreme Environments

TIMELINE OF THE PROJECT

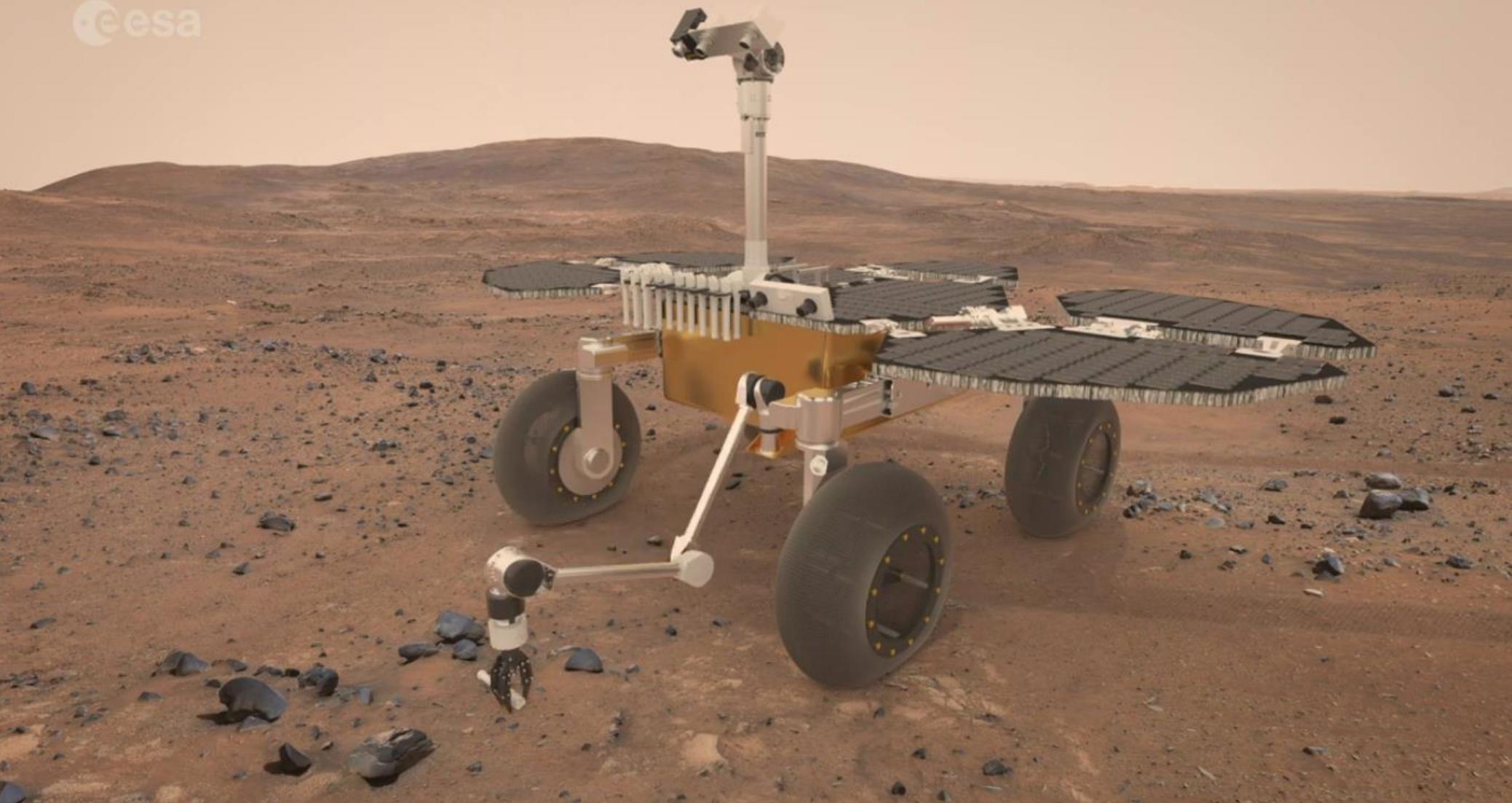


File D.E.M Visual Robot Perception Localization Locomotion Energy Soil

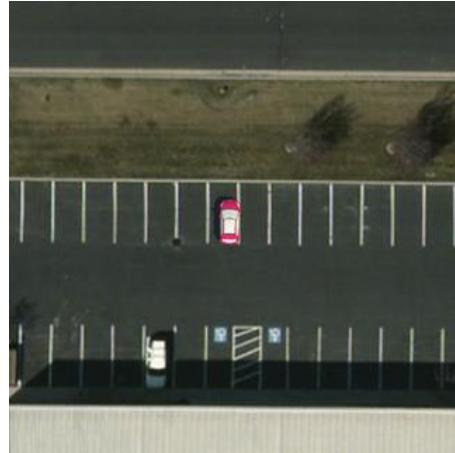


Sample Fetch Rover (SFR)

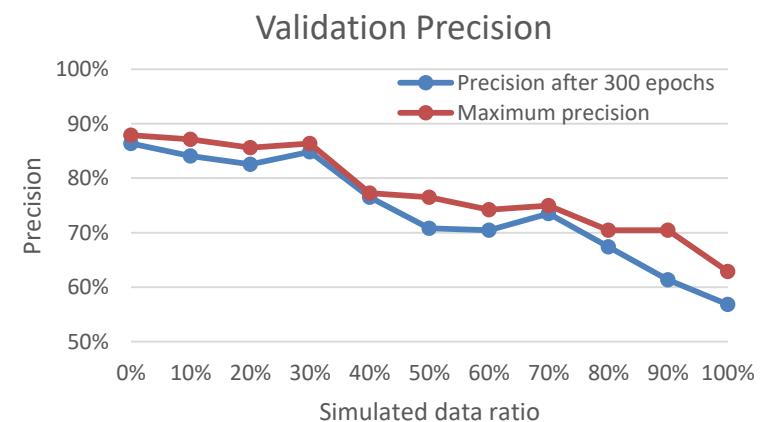
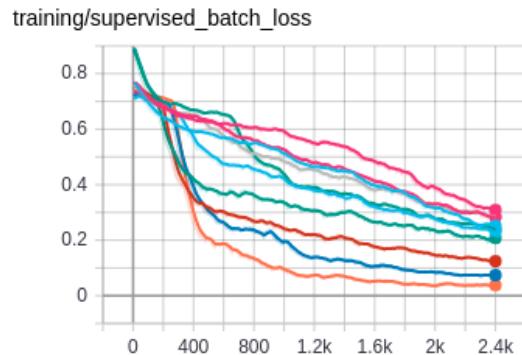
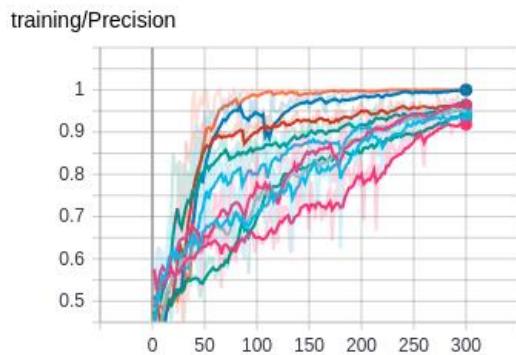
esa



Apprentissage d'un réseau de neurones à partir de données simulées



Analyse des métriques d'apprentissage et de validation



Introduction

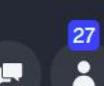
- I. Magellium
- II. Métier et Projets
- III. Computer Vision

- Qu'est-ce que la vision par ordinateur?



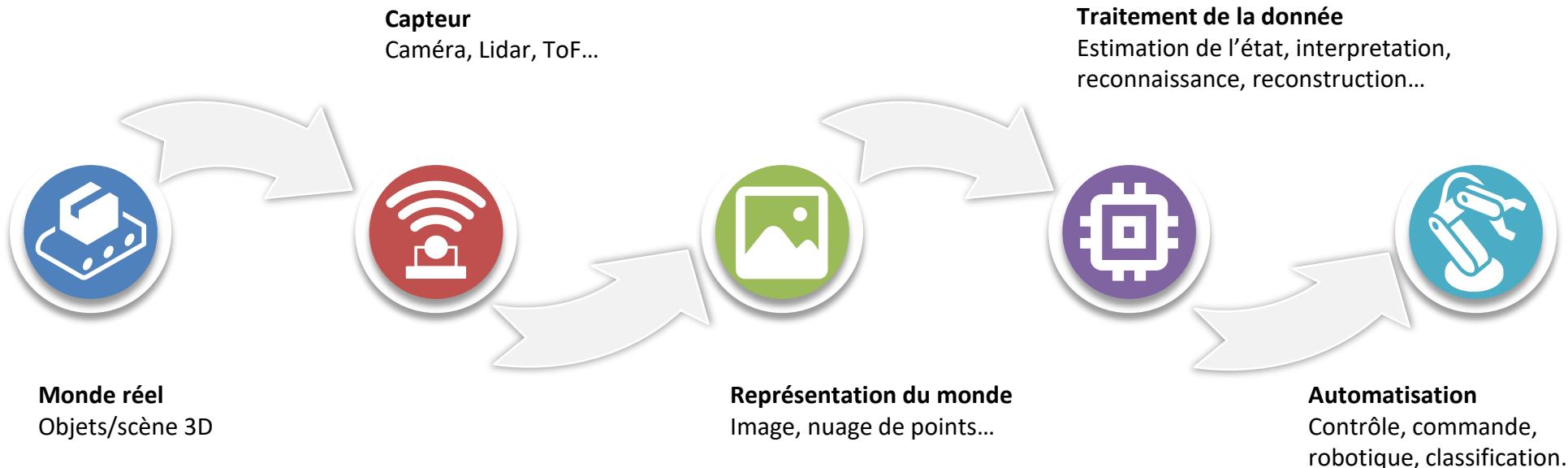
www.menti.com code 9409 5594

Qu'est-ce que la vision par ordinateur?



La **vision par ordinateur** est un domaine interdisciplinaire dont le principal but est de permettre à une machine d'**analyser**, **traiter** et **comprendre** une **représentation** de l'environnement obtenue par un système d'**acquisition**.

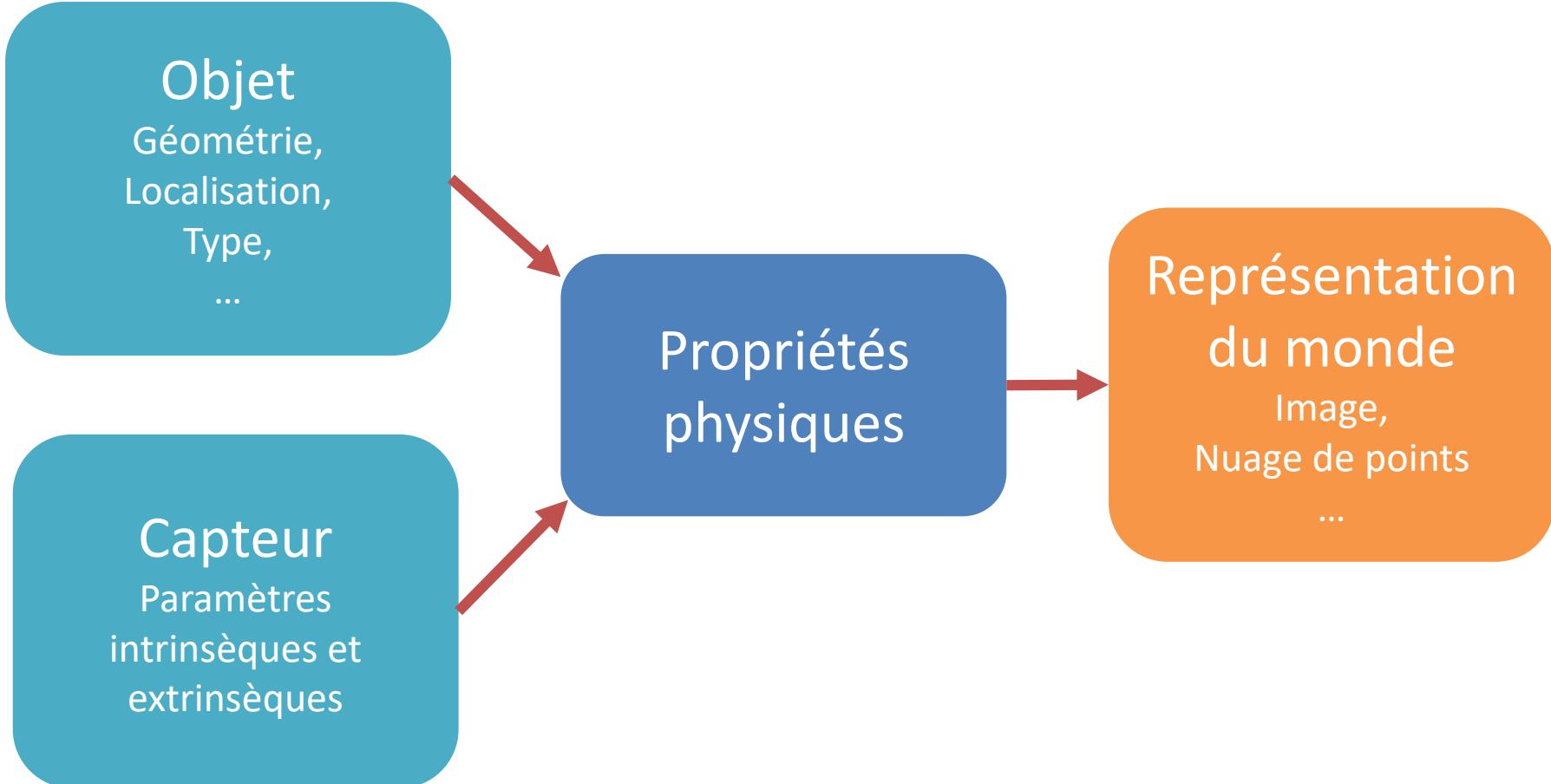
Cette branche de l'intelligence artificielle implique le développement d'**algorithmes** permettant l'**automatisation** de tâches que le système visuel humain peut réaliser.

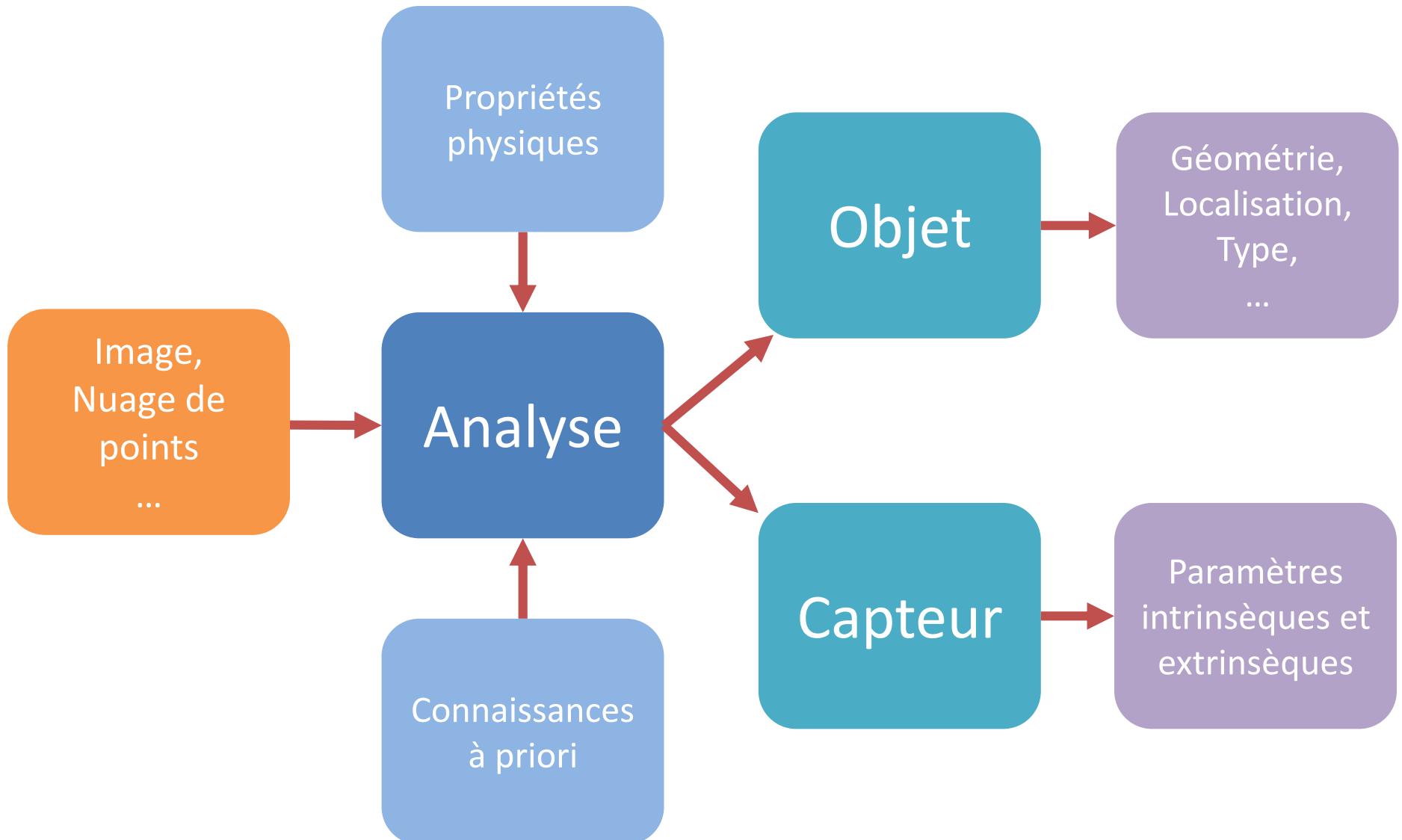


Partie I: Perception 3D

- I. Introduction
- II. Systèmes d'acquisition 3D
- III. Modélisation de la camera
- IV. Stéréovision

- Qu'est-ce que la perception?
 - **Estimation** des propriétés **géométriques** et **sémantiques** d'un ou plusieurs objets, à partir d'**observations** produites par un **système d'acquisition**.
- Que peut-on obtenir?
 - Position et/ou orientation du capteur
 - Localisation des objets dans une scène
 - Reconstruction 3D d'objets ou d'une scène
 - Reconnaissance d'objets
 - Interprétation sémantique
 - Interprétation d'images
 - ...

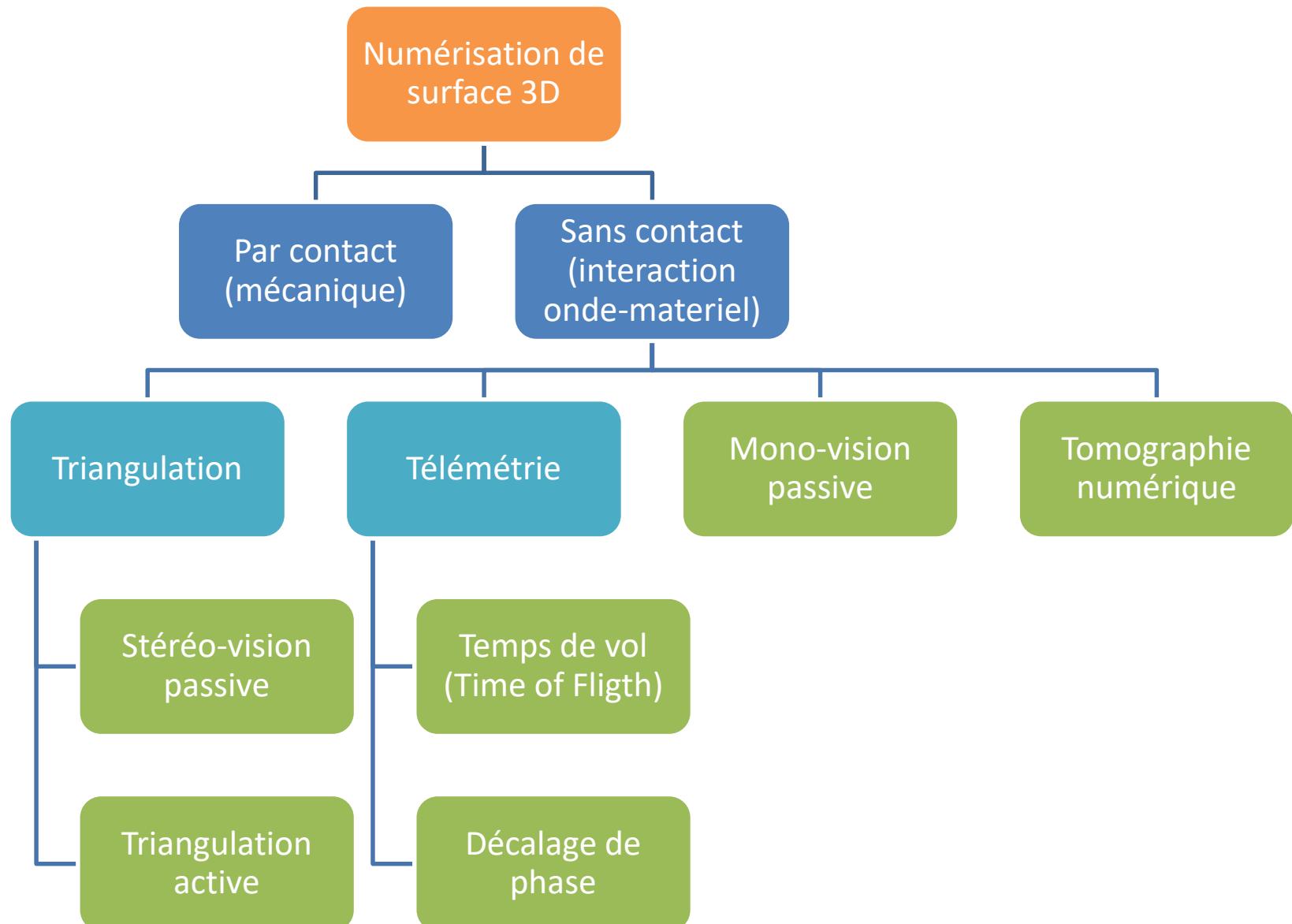




Partie I: Perception 3D

- I. Introduction
- II. Systèmes d'acquisition 3D
- III. Modélisation de la camera
- IV. Stéréovision

Systèmes d'acquisition 3D



Numérisation par contact

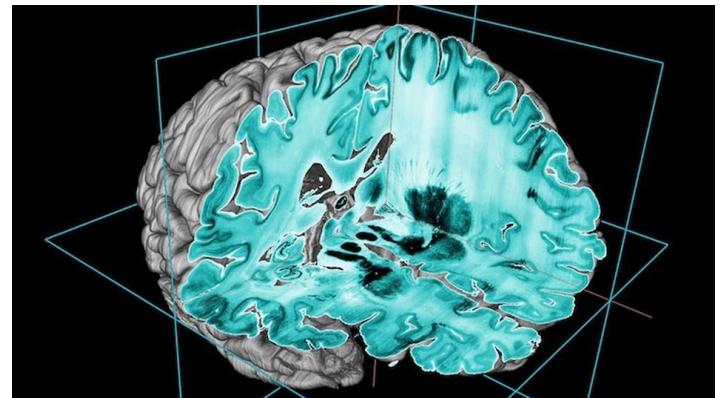
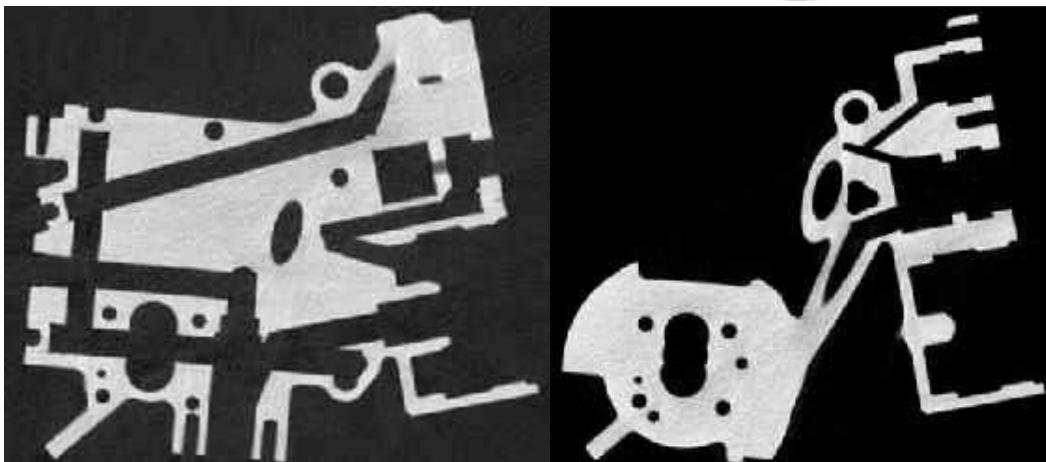
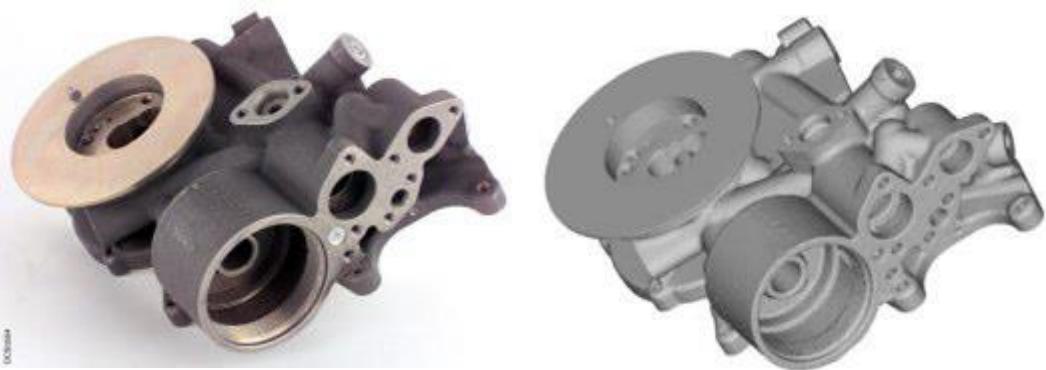


Machine à mesurer tridimensionnelle

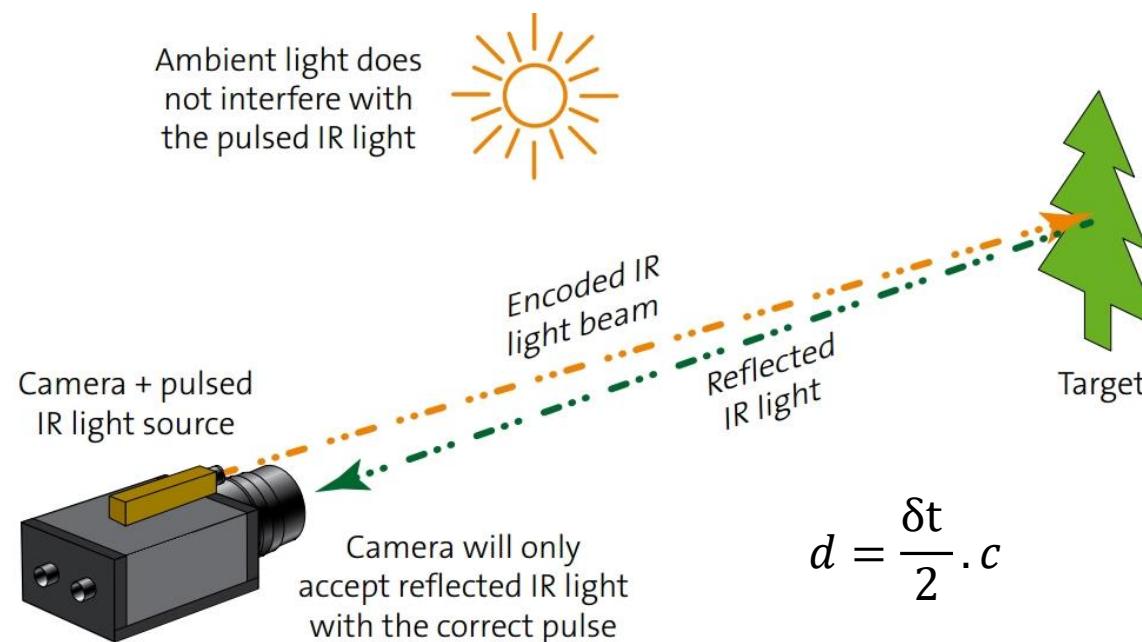


Bras articulé

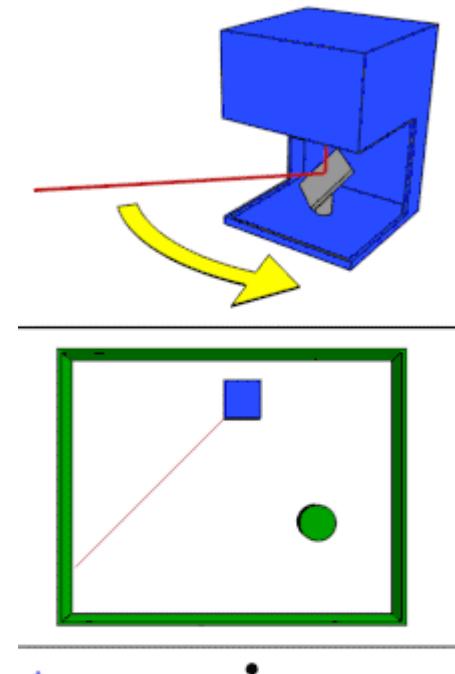
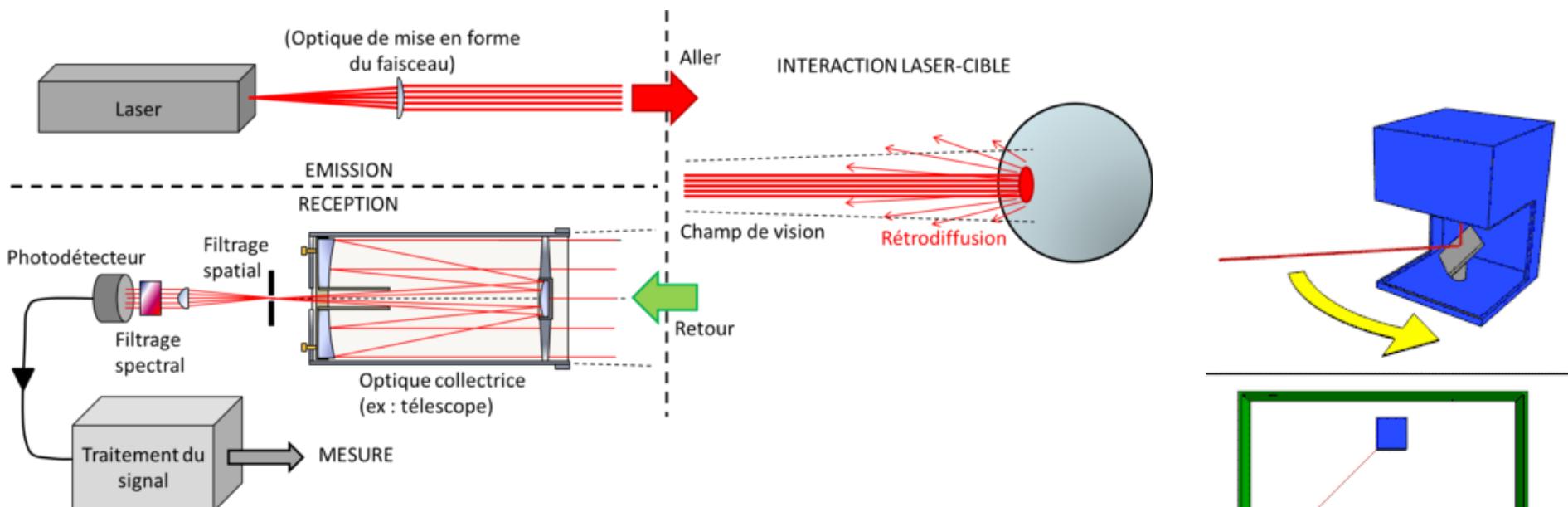
- Rayons X
 - Principe de l'IRM
 - Superposition de coupes 2D successives



- Envoi d'une impulsion pulsée:
 - Lumineuse classique: Caméra ToF
 - Lumineuse Laser : LIDAR
 - Ultrasonore : SONAR
 - Ondes radio : RADAR
- Mesure du temps de vol (ToF: Time of Flight)

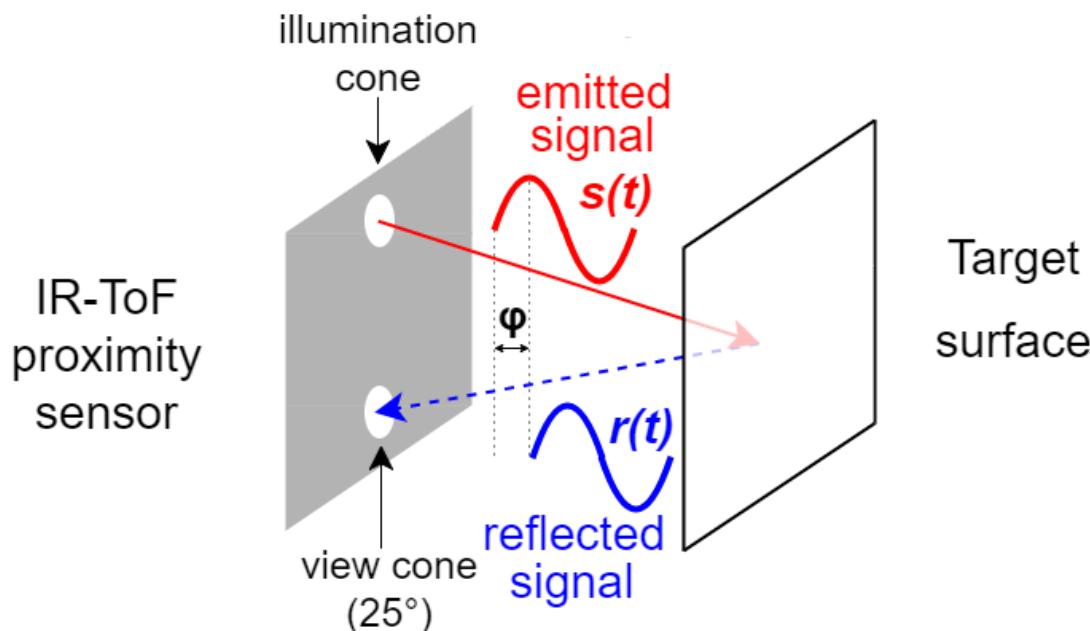


Lidar ToF



Télémétrie par décalage de phase

- Envoi d'une impulsion modulée:
 - Lumineuse classique : Caméra SWIR à modulation de phase
 - Lumineuse Laser : LIDAR
 - Ultrasonore : SONAR
 - Ondes radio : RADAR
- Mesure du décalage de la phase

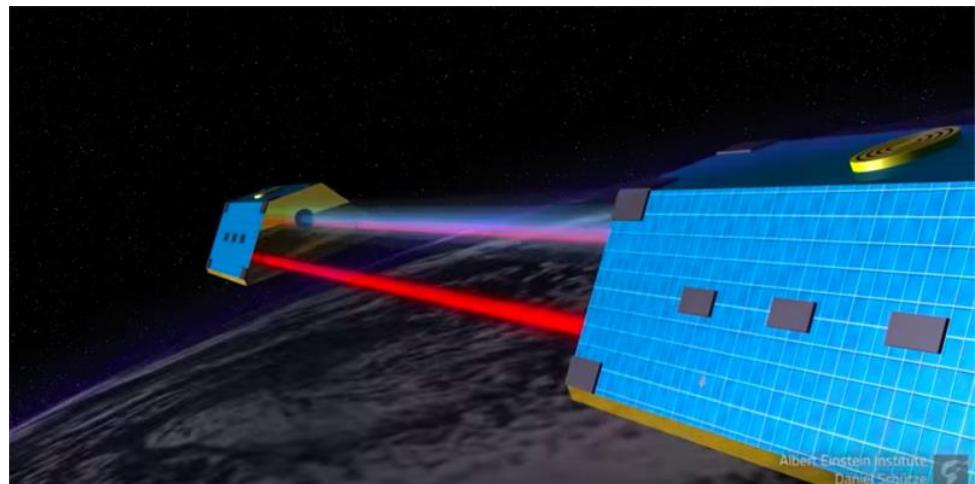
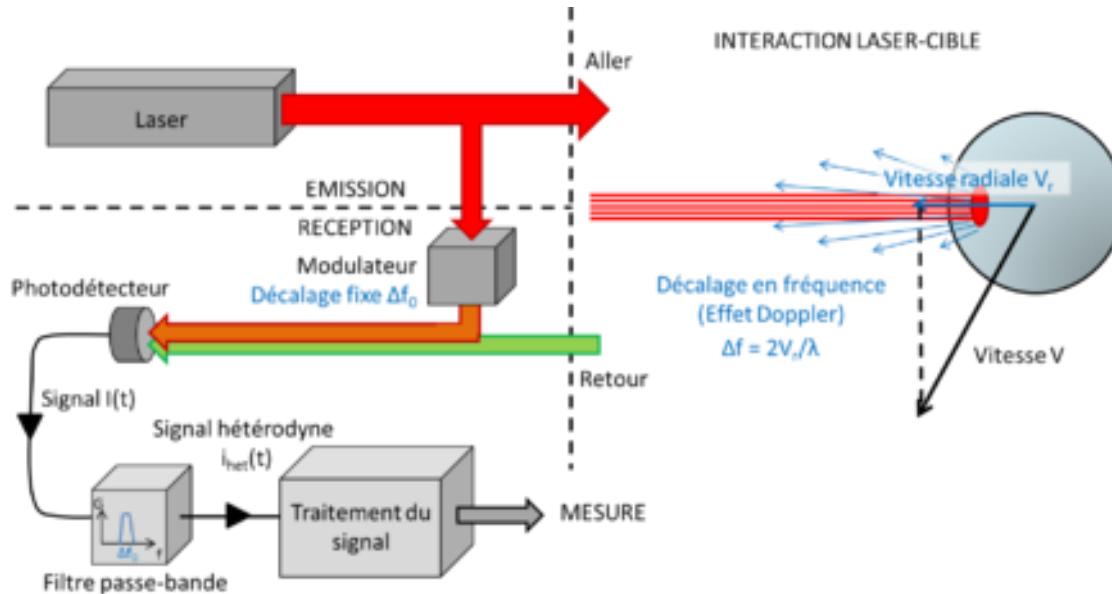


$$\varphi = \delta t \cdot 2\pi \cdot f$$

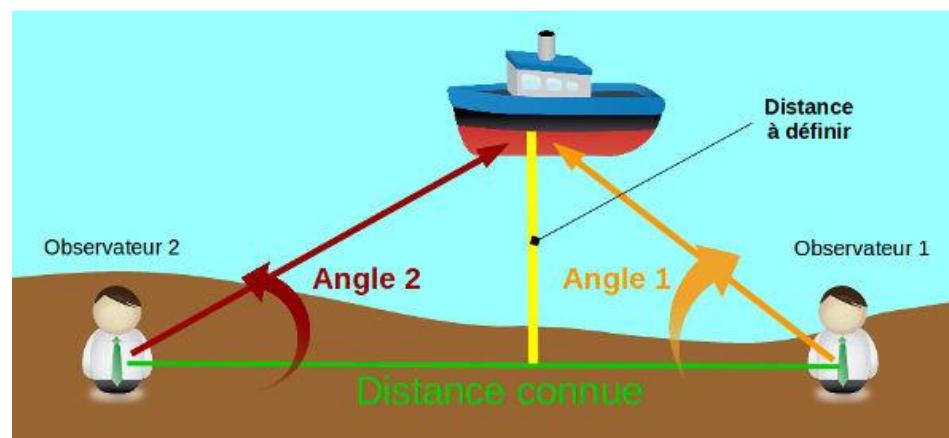
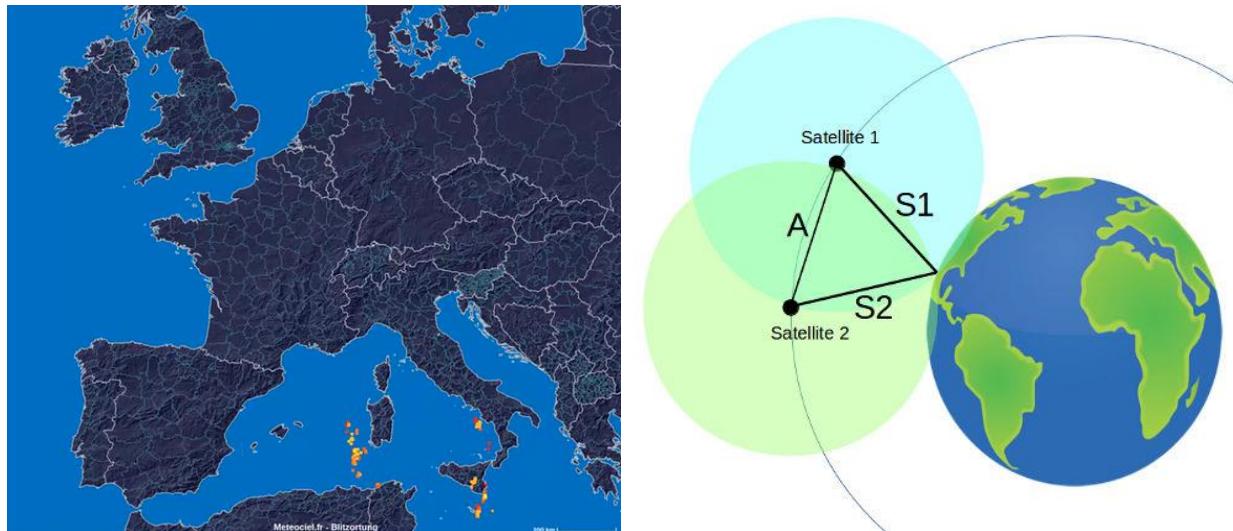
$$\text{et } d = \frac{\delta t}{2} \cdot c$$

$$\text{donc } d = \frac{\varphi}{2\pi} \cdot \frac{c}{2f}$$

Télémétrie par décalage de phase

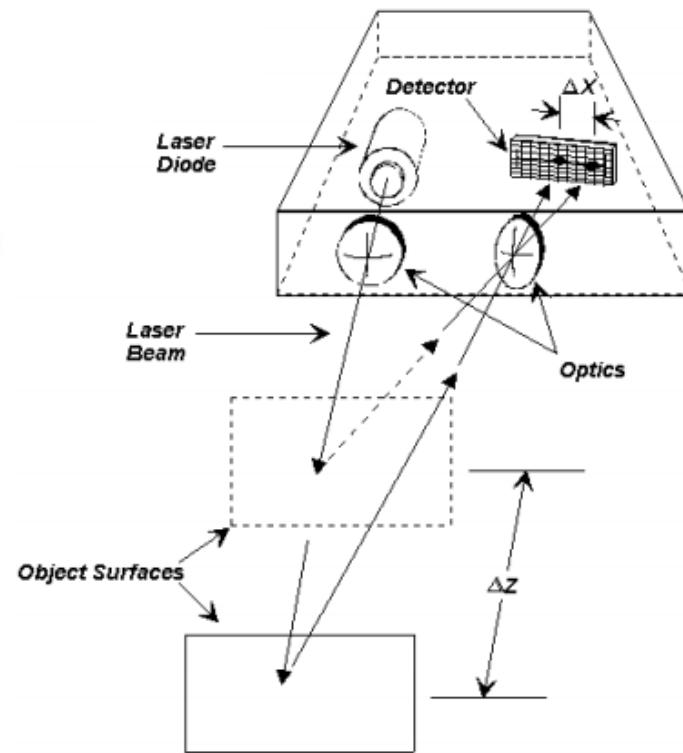
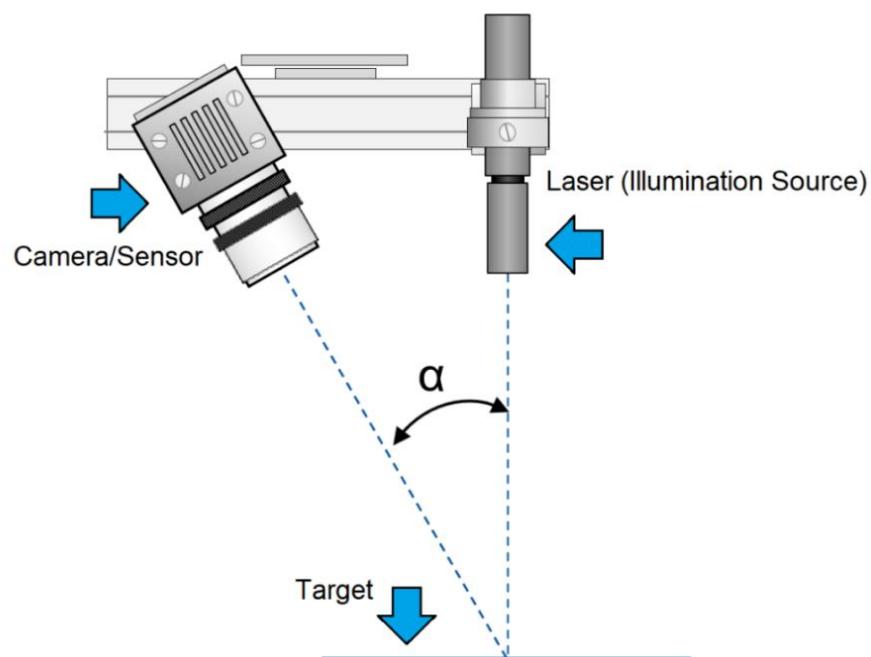


- Principe utilisé par les géomètres, le GPS...

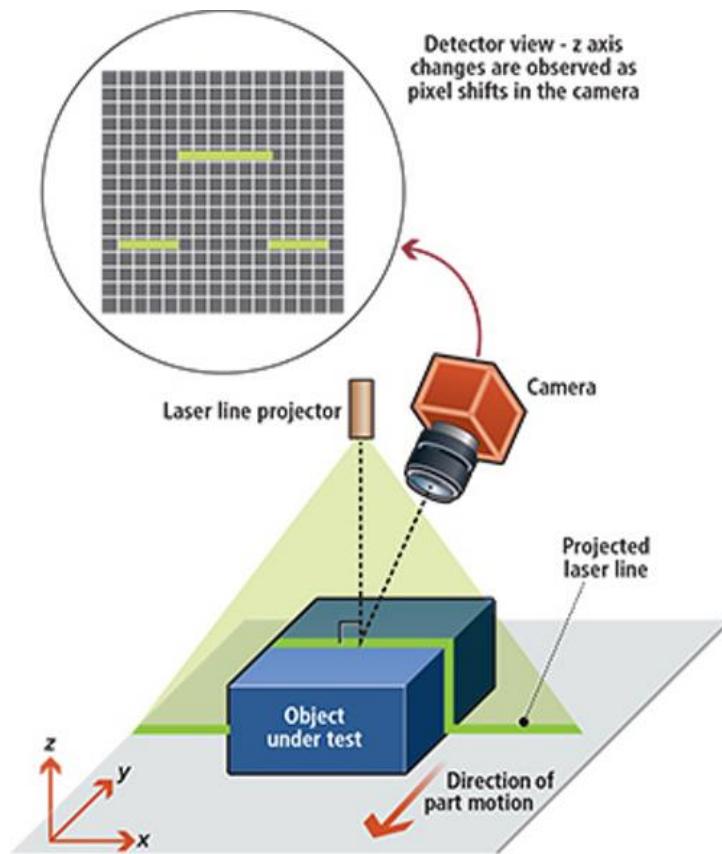


- Pointeur laser

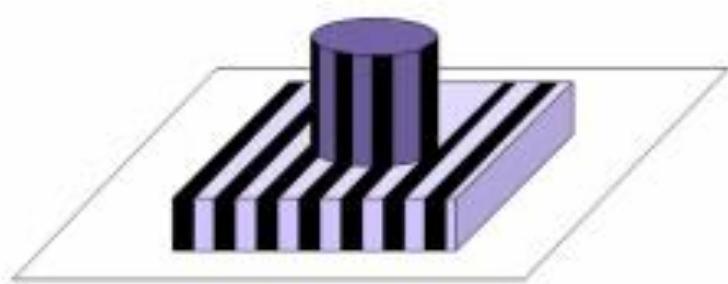
- Un rayon laser est envoyé vers l'objet à mesurer
- La lumière diffusée est observée par une caméra
- On peut en déduire la profondeur du point



- Laser ligne, profilométrie
 - Une image saisie donne une ligne de points
 - Un unique balayage suffit pour assurer la couverture de la surface



- Lumière structurée
 - Projection d'un motif connu
 - Appariement pixels/motif



- Avantages:

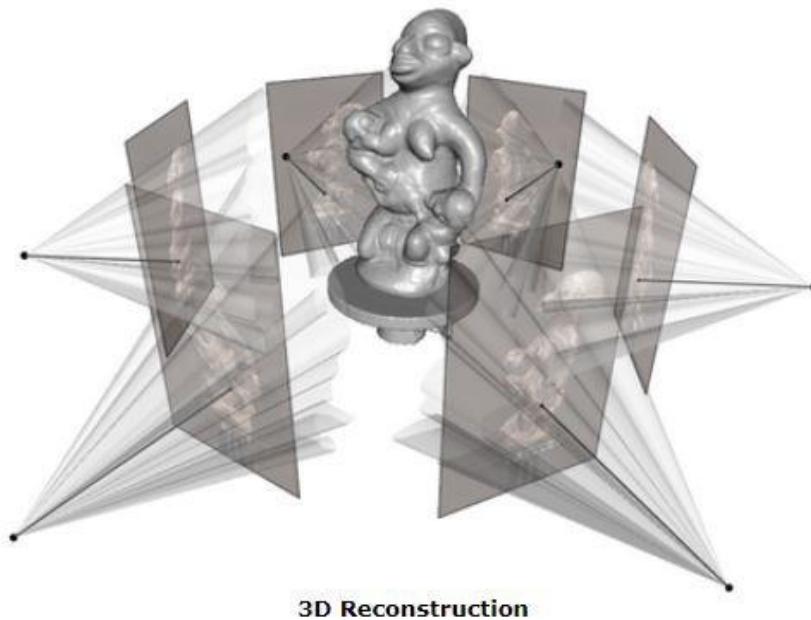
- Pas de contact
- Peu onéreux
- Robuste (~pas de pièce mobile, pas d'interférences...)
- Facilité d'acquisition de grandes quantités de données
- Couverture dense
- Grande variété de distances
- Possibilité d'obtenir de l'information 2D et 3D
- Pas besoin d'éclairage spécifique: technique passive
- Information géométrique mais aussi sémantique, interprétation de l'image...
- Donnée directement interprétable par l'Homme

- Inconvénients:

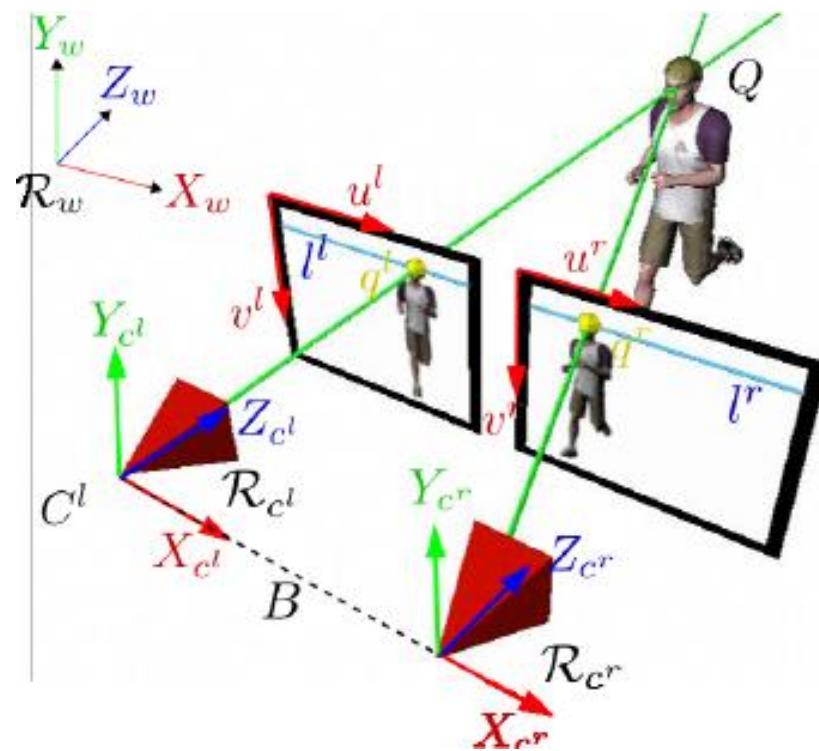
- Besoin de lumière
- Occlusions
- Perte d'information: une image est une projection du monde 3D sur un plan 2D
- Difficulté de l'appariement des pixels
- Précision relativement faible



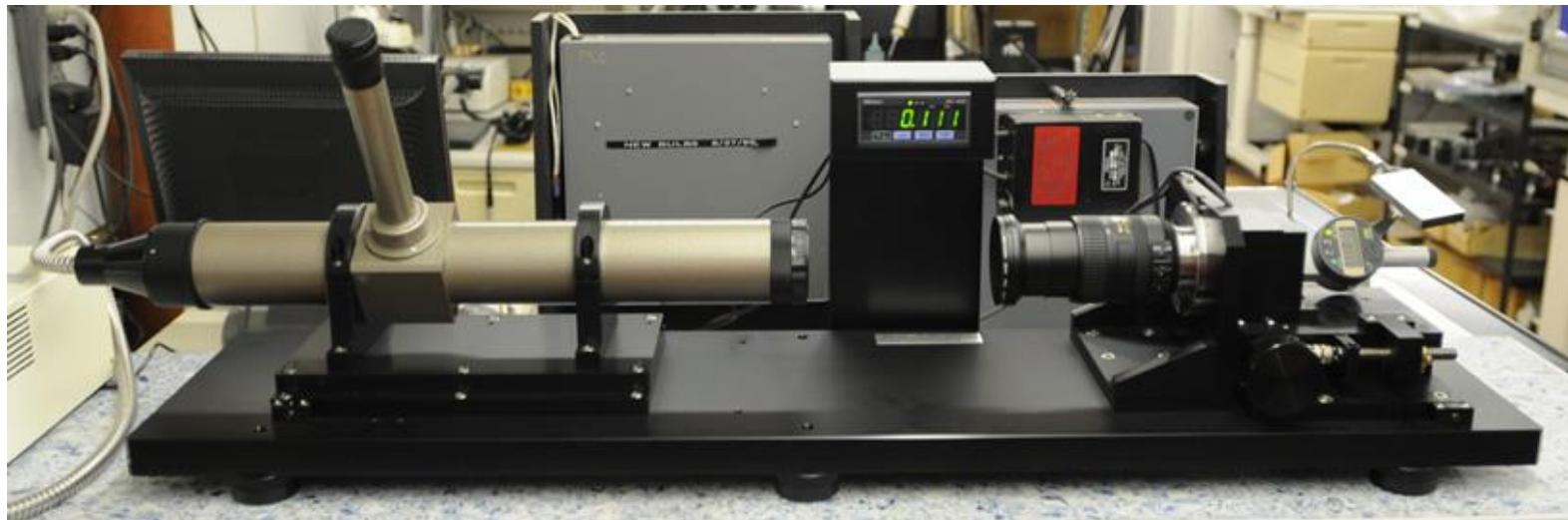
- Utilisation de simples images 2D pour obtenir de l'informations 3D
- Obtention d'angles de vue différents par mouvement de l'objet et/ou de la caméra



- Vision d'une même scène depuis deux endroits légèrement décalés l'un par rapport à l'autre
- Principe de la perception 3D chez l'Homme



- Les points 3D sont des mesures géométriques obtenues par des principes physiques (lumière, contact, etc.) et mécaniques
- Les erreurs systématiques de mesure peuvent être améliorées par calibrage/étalonnage



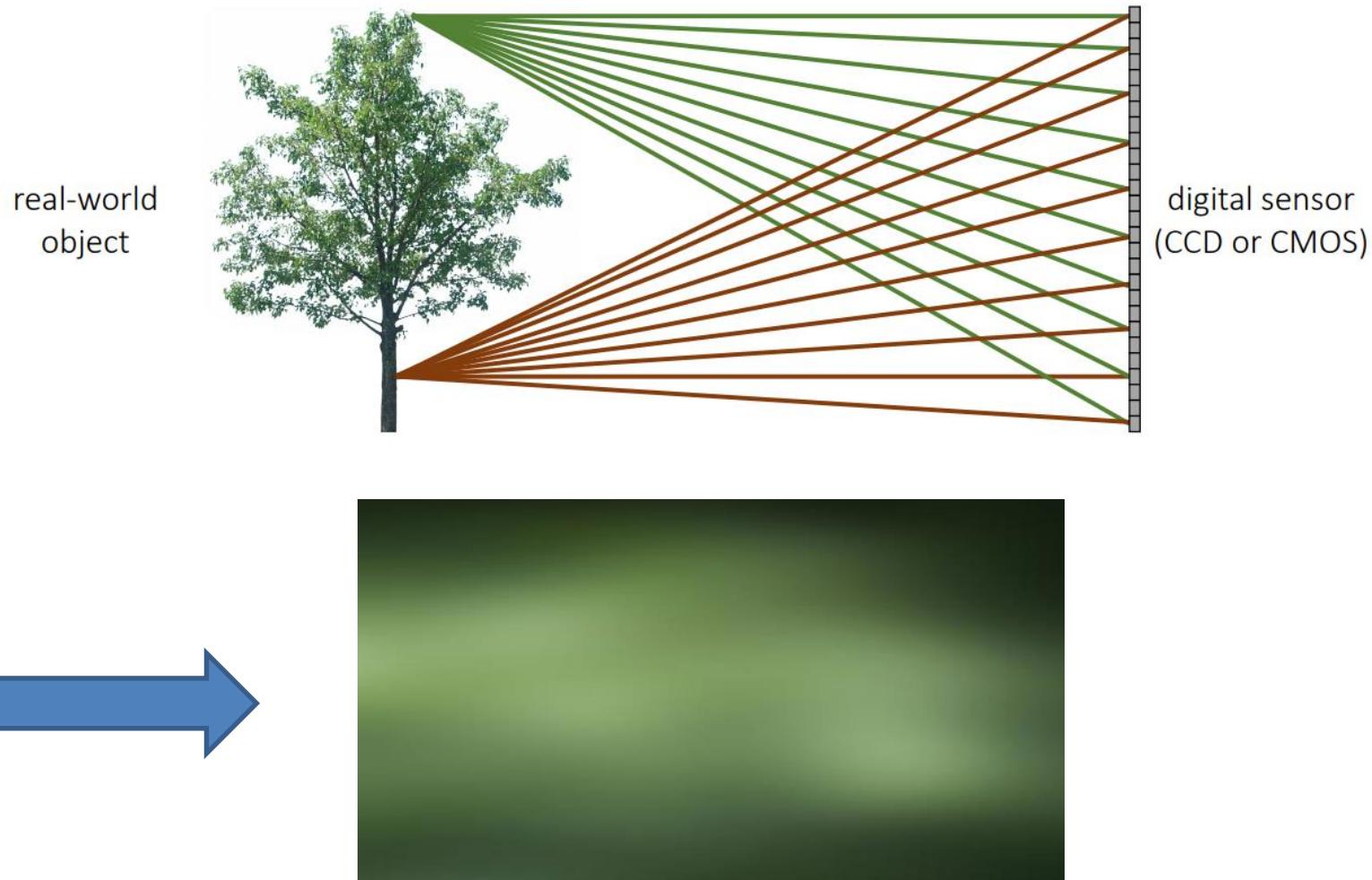
Partie I: Perception 3D

- I. Introduction
- II. Systèmes d'acquisition 3D
- III. Modélisation de la camera
- IV. Stéréovision

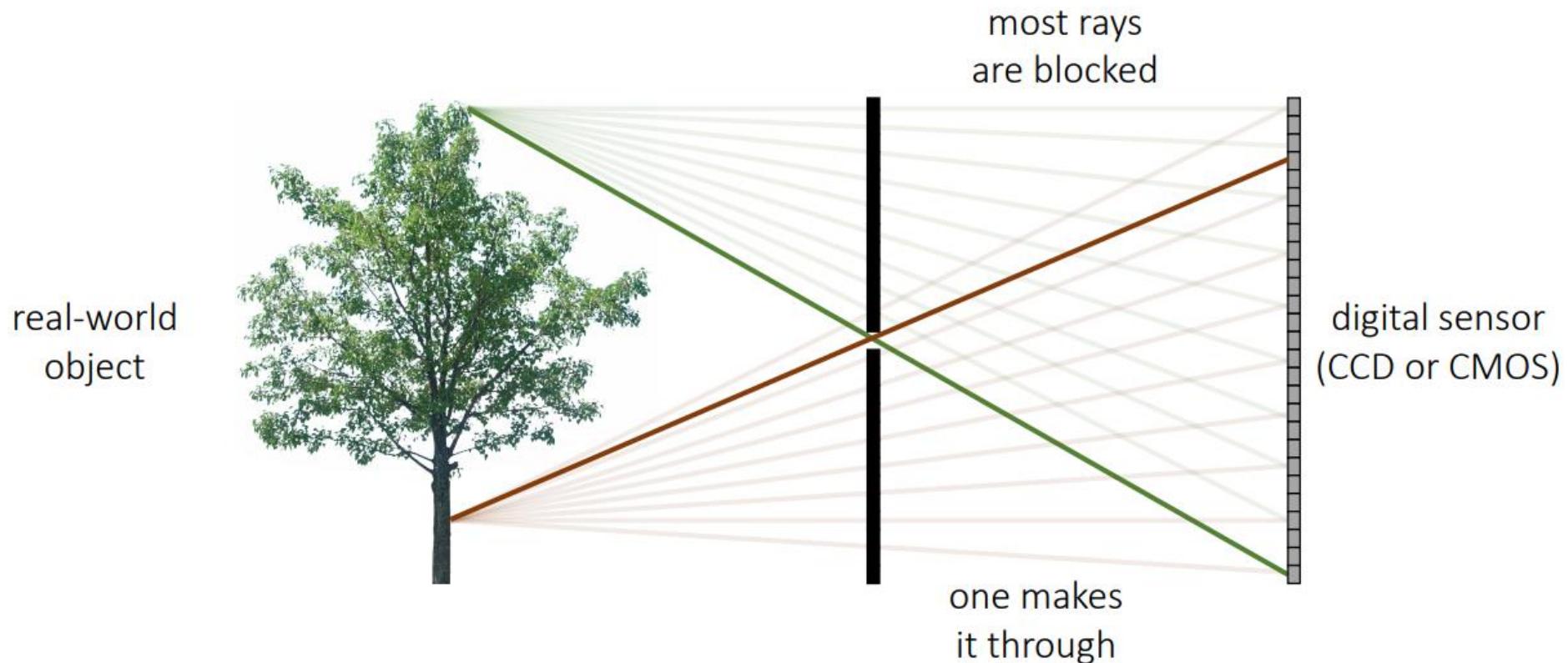
- Postulats:

- La propagation de la lumière est décrite par des rayons lumineux provenant d'une source de lumière
- Un rayon lumineux suis une ligne droite dans un milieu homogène
- À l'intersection entre deux milieux homogènes, la lumière est réfléchie ou réfractée
- Le chemin parcouru par un rayon lumineux est réversible
- L'intersection de rayons lumineux est sans effet sur ces même rayons

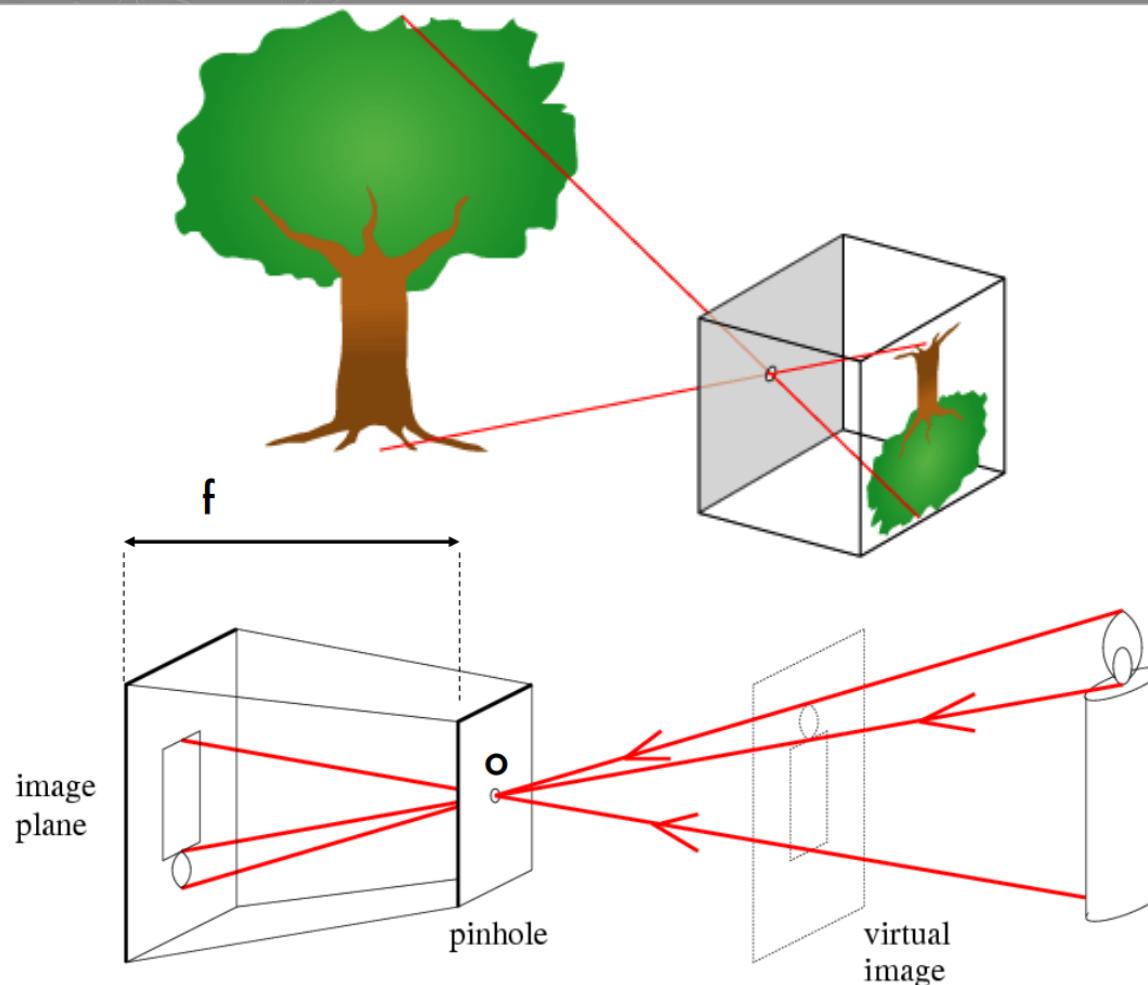
- Capteur uniquement



- Pinhole/sténopé



Modèle Sténopé

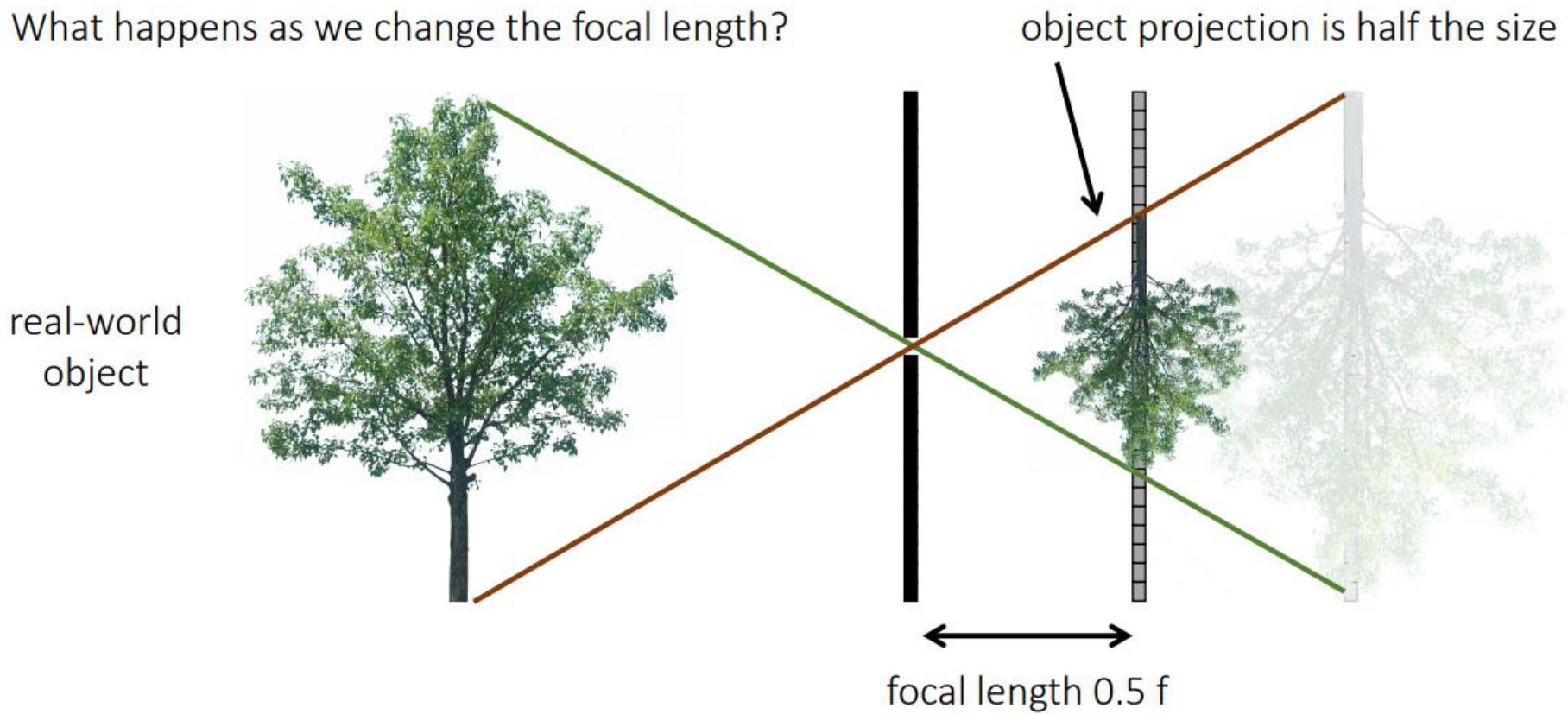


f = focal length

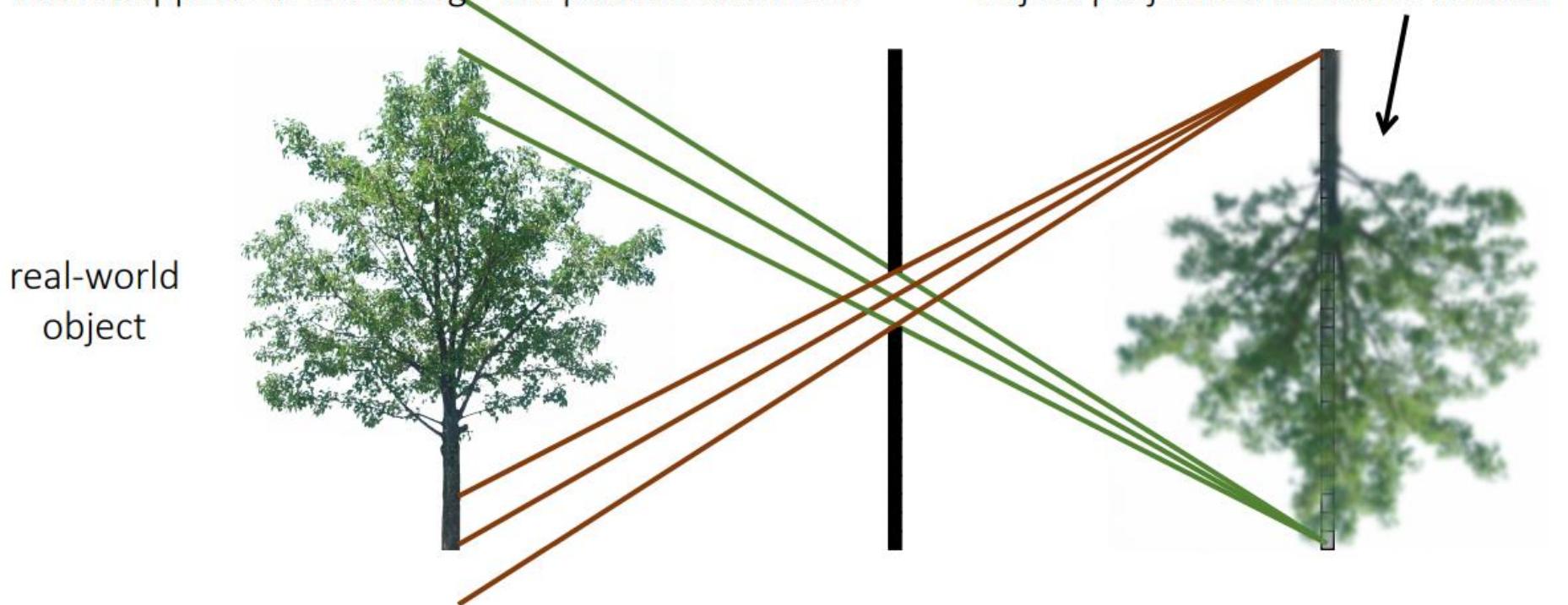
o = aperture = pinhole = center of the camera

Distance focale

What happens as we change the focal length?



What happens as we change the pinhole diameter?



- Réduction de la taille de l'ouverture
 - Amélioration de la netteté
 - Réduction de la luminosité



f / 1.4



f / 2.8

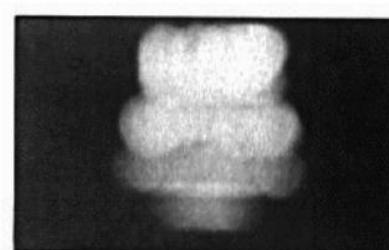
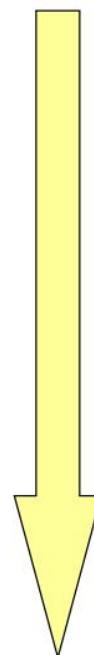
Shrinking
aperture
size



f / 4



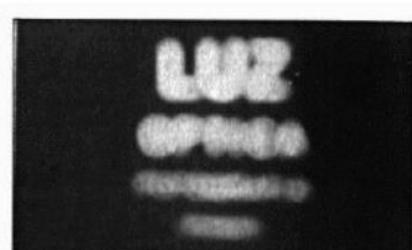
f / 8



2 mm



0.6mm



1 mm

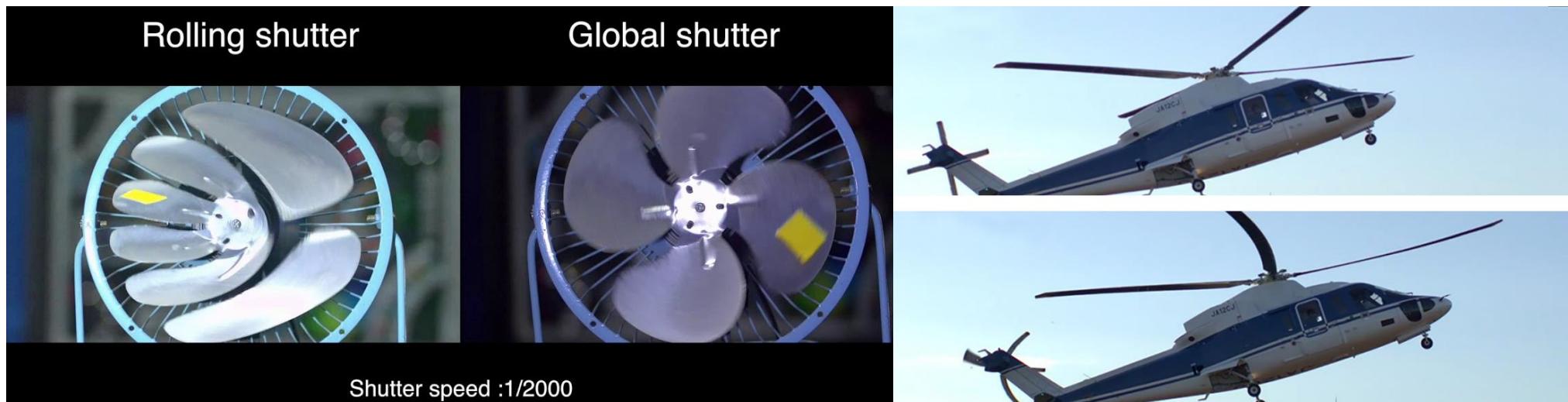


0.35 mm

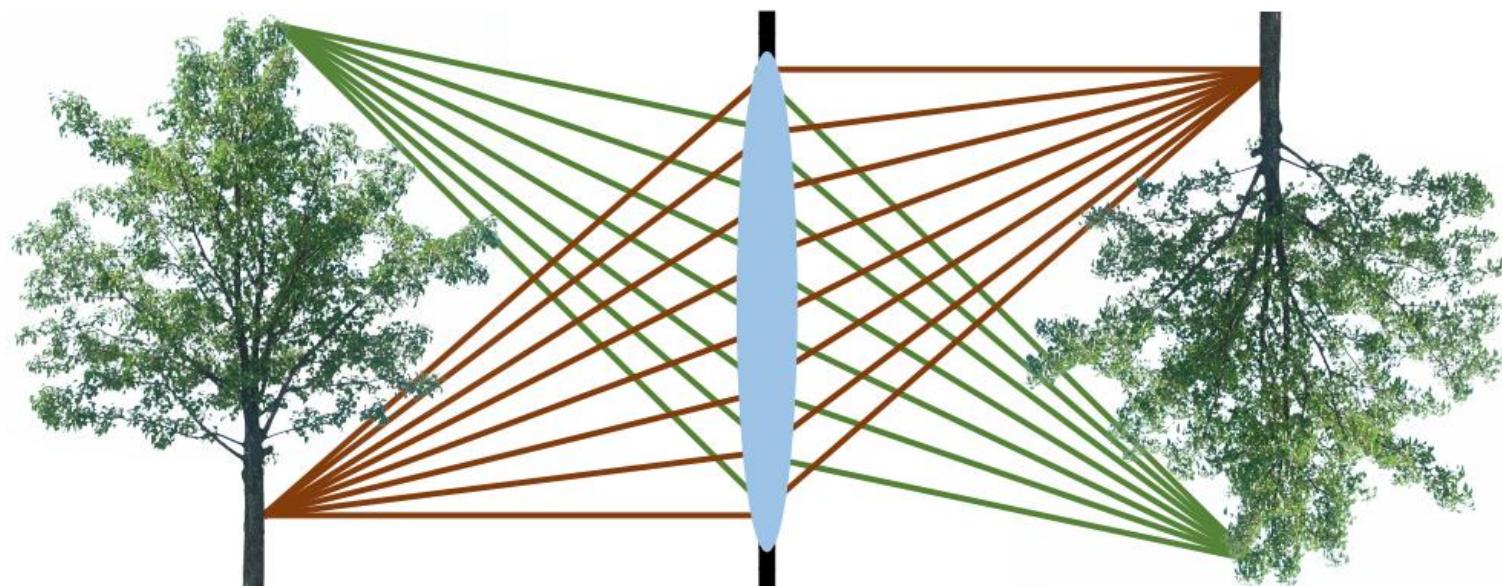
- Sans lumière, pas d'image
- L'illumination de la scène a une influence importante sur le processus d'acquisition
- Contrôler l'illumination est un concept clef
- Globalement possible de contrôler l'illumination dans l'industrie
- Difficile à impossible en milieu extérieur



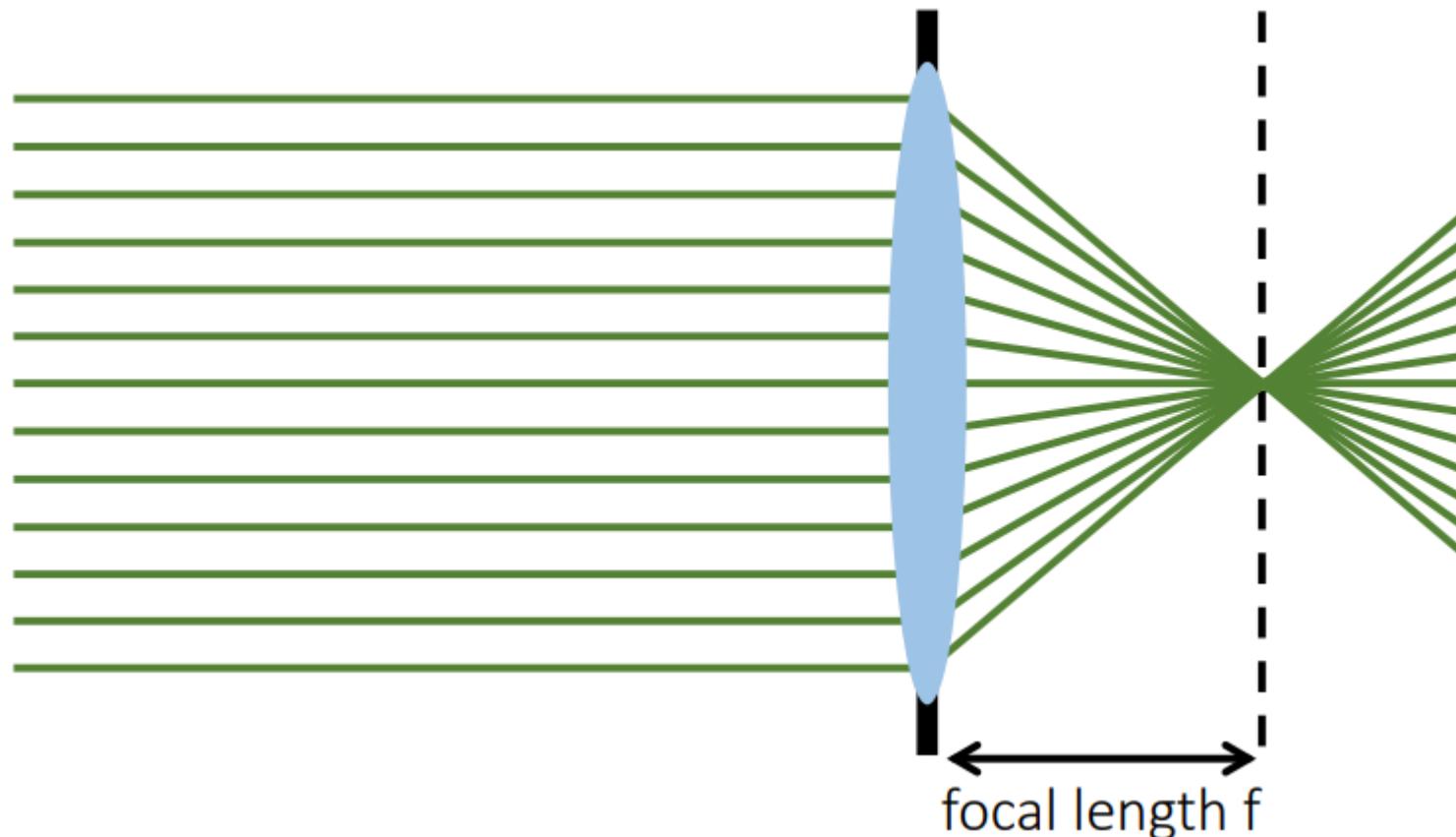
- Il faut si possible contrôler l'éclairage
- Sinon, agir sur les paramètres physiques de la caméra
 - Vitesse d'obturation
 - Ouverture
- Petite vitesse d'obturation et/ou grande vitesse de l'objet
=> attention au flou de mouvement
- Logiciel d'auto-exposition
- Caméras « global shutter » plutôt que « rolling shutter »



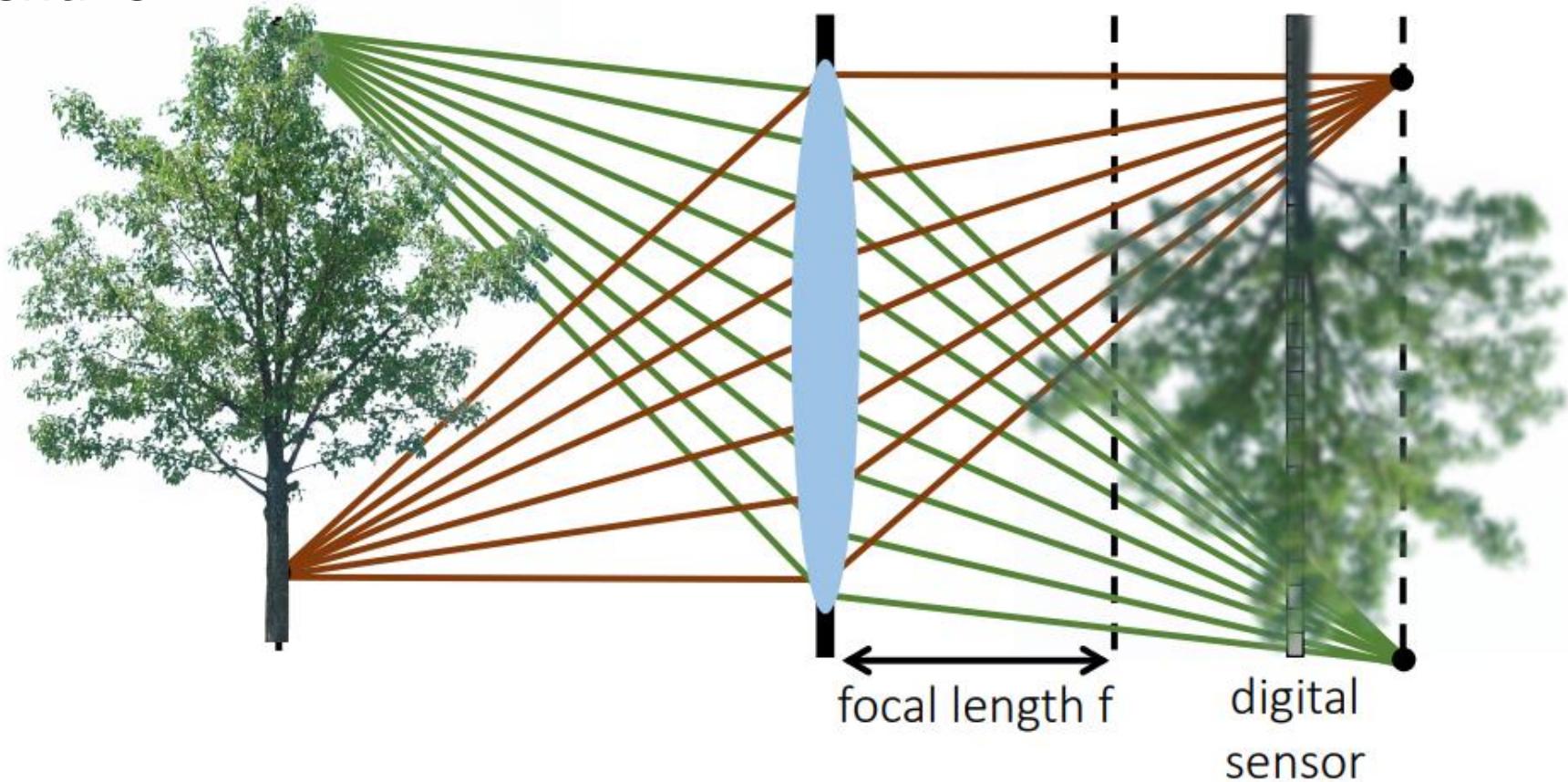
- Utilisation d'une lentille pour concentrer la lumière sur le plan image
 - Amélioration de la luminosité
 - Amélioration de la netteté
- Le choix de la lentille dépend
 - De la distance de travail
 - Du champ de vision recherché



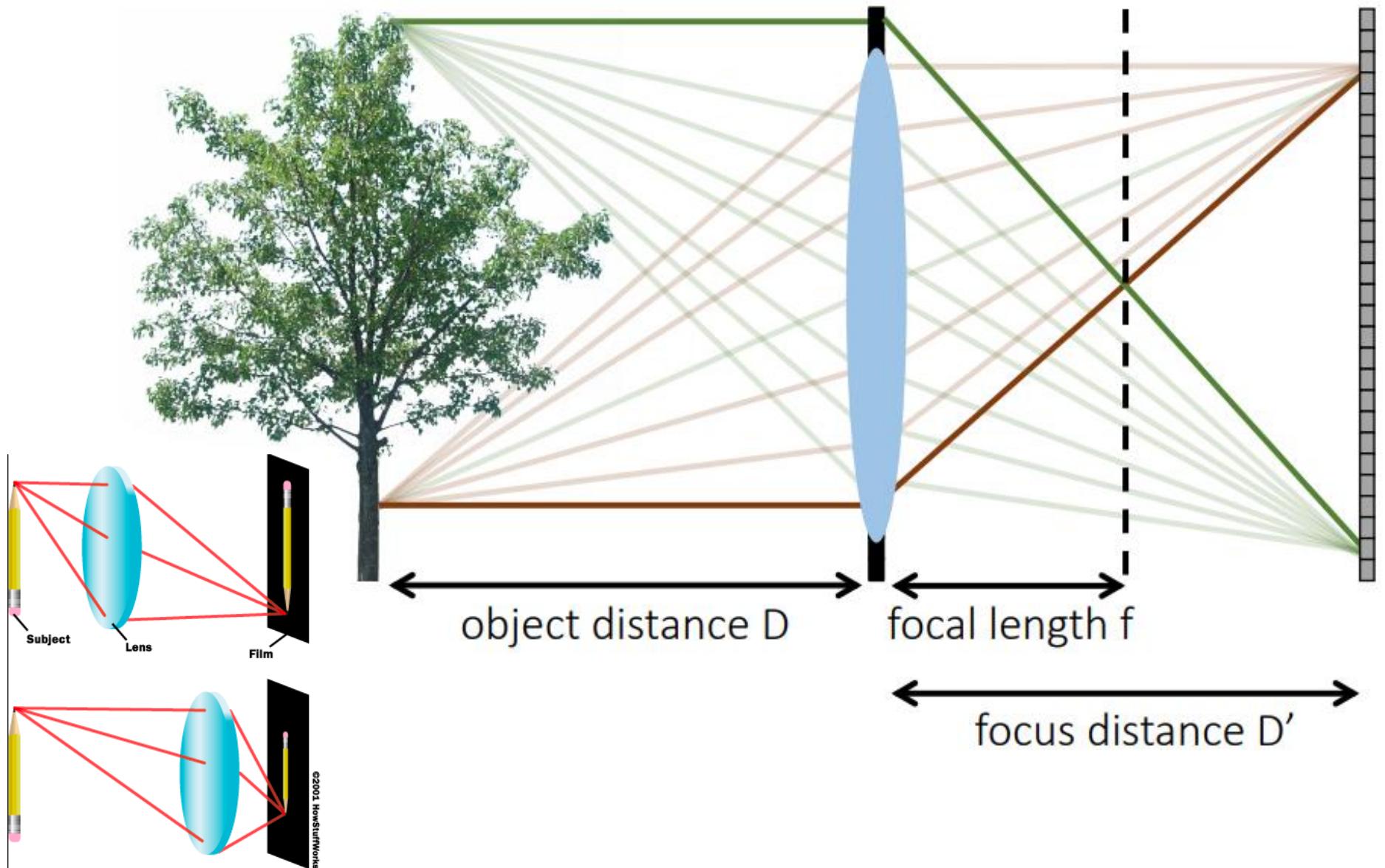
- Tous les rayons parallèles à l'axe optique convergent en un point unique, le point focal, sur un plan situé à la distance focale du centre de la lentille



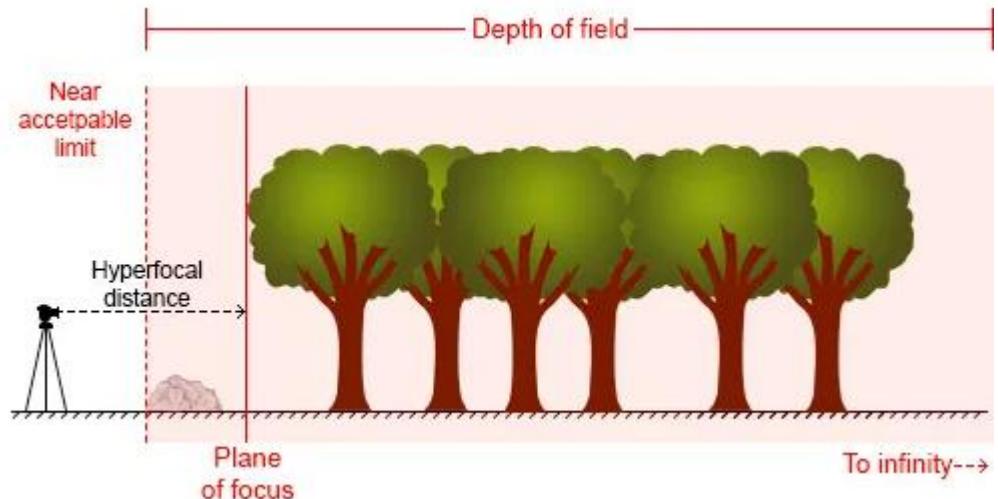
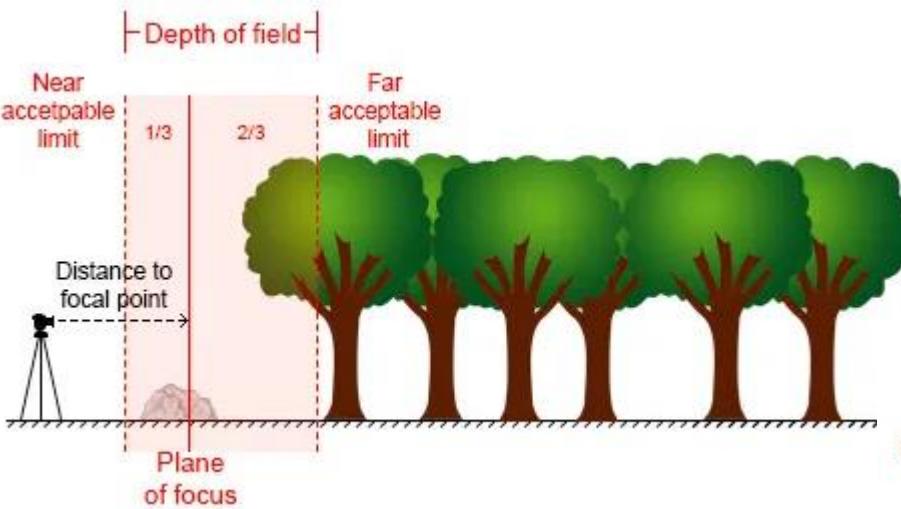
- Les objets sont correctement projetés sur le plan image uniquement lorsqu'ils sont à une certaine distance de la lentille/lorsque le capteur est à une certaine distance de la lentille



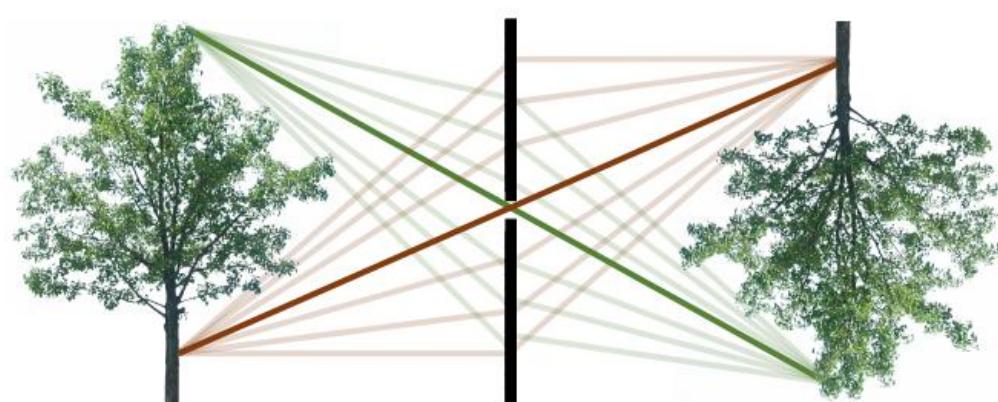
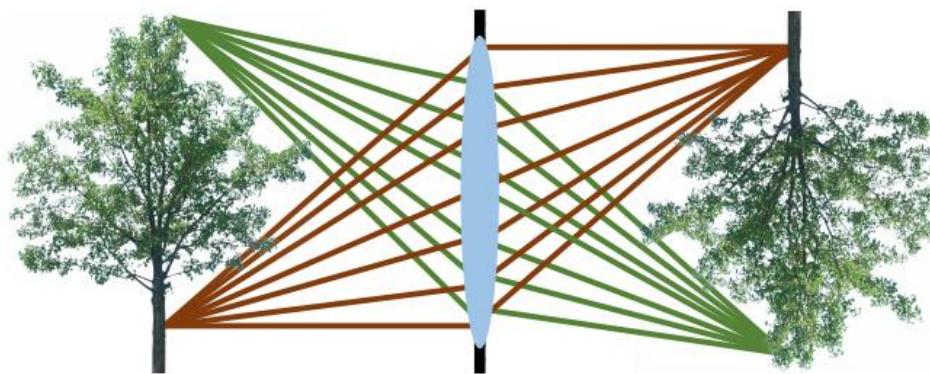
Distance de focus



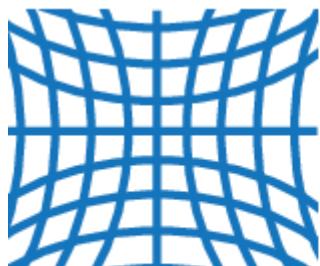
- Distance minimum à laquelle il est possible de faire la mise au point tout en gardant les objets situés à l'infini avec une netteté acceptable.
- Dépends de la distance focale et de l'ouverture



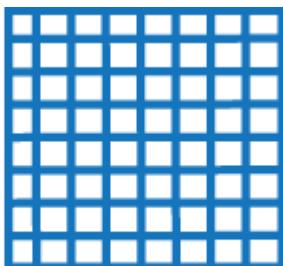
- On peut assimiler une caméra munie d'une lentille à une caméra sténopé si:
 - On considère uniquement le rayon lumineux central
 - On suppose que la mise au point est faite
 - On considère que la distance de focus de la caméra avec lentille est équivalente à la distance focale de la caméra sténopé
 - On néglige ou corrige les distorsions induites par la lentille
 - On ajoute un système d'ouverture pour limiter les erreurs



- Problème des lentilles
 - Distorsion radiale
 - La distorsion est plus importante pour les rayons qui passent près du bord de la lentille



Negative radial distortion
"pincushion"



No distortion

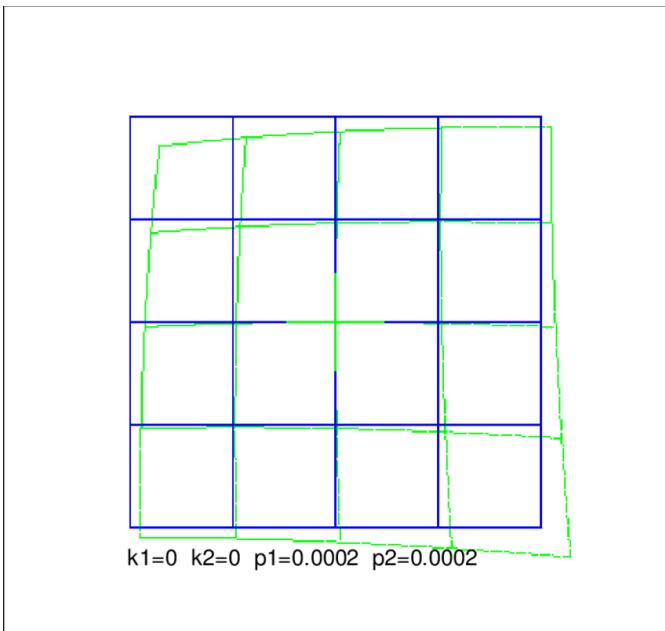
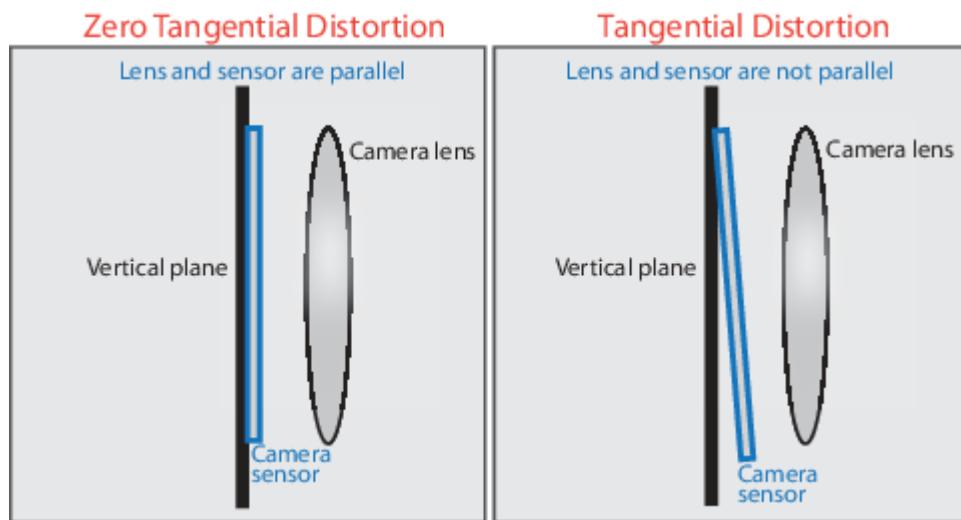


Positive radial distortion
"barrel"

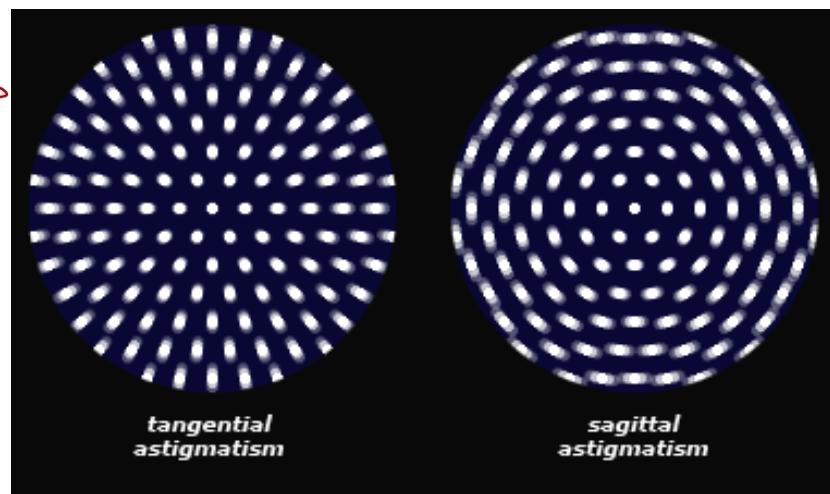
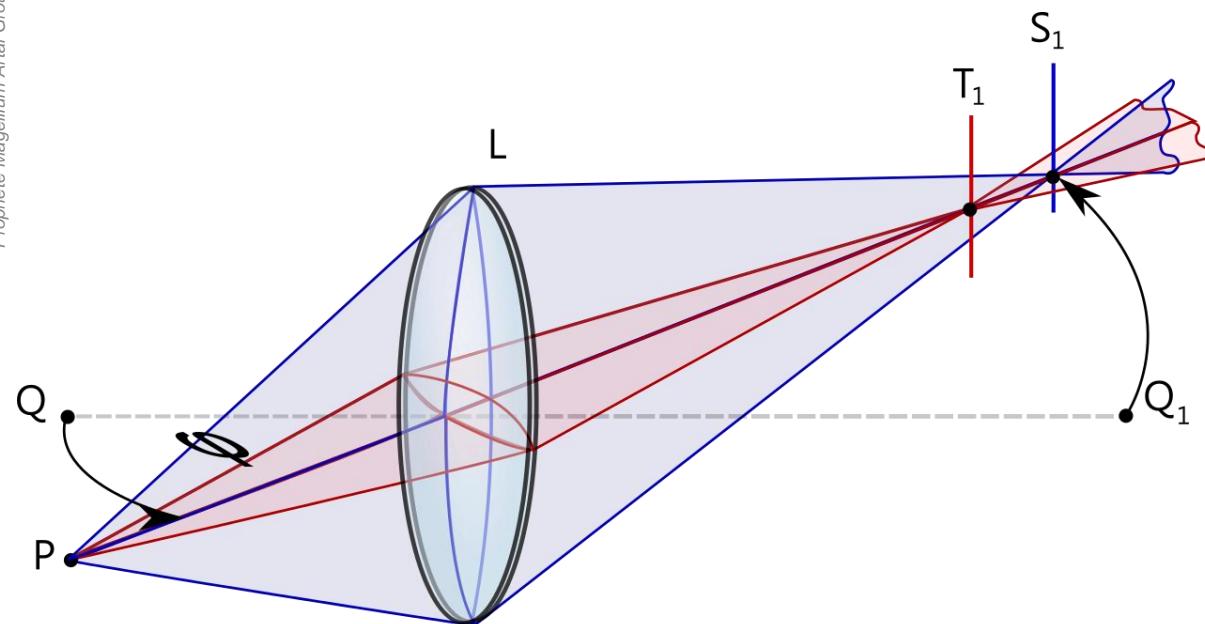


Image magnification decreases
with distance from the optical axis

- Problème des lentilles
 - Distorsion tangentielle
 - La distorsion est plus importante pour les rayons qui passent près du bord de la lentille

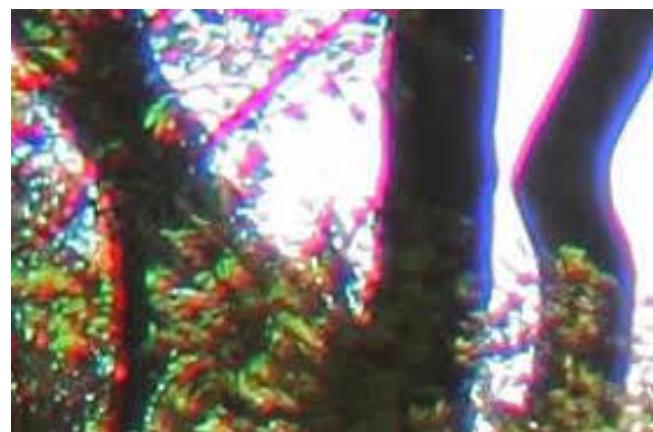
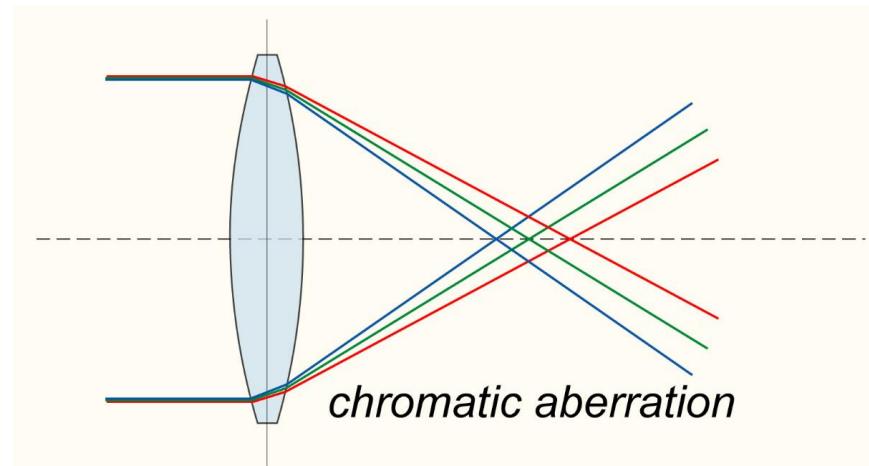
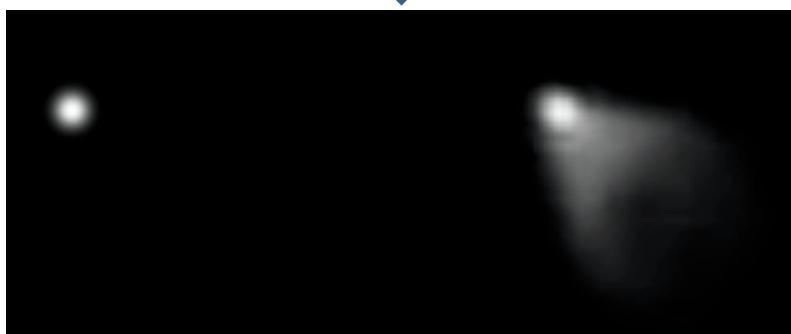
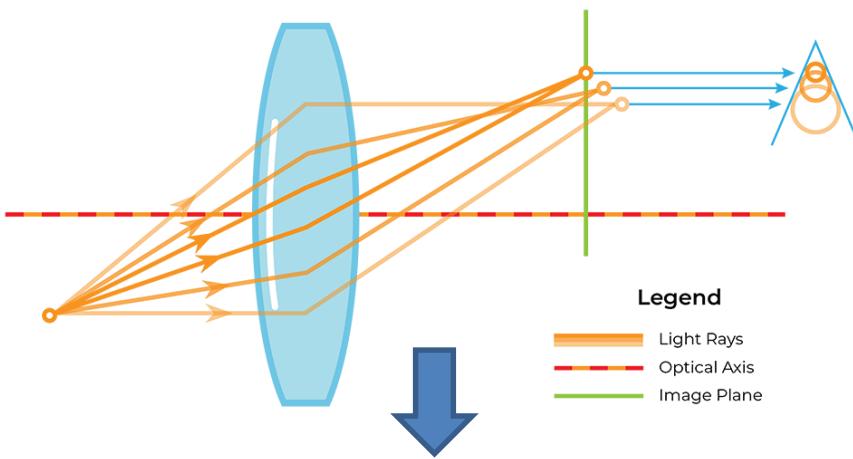


- Problème des lentilles
 - Astigmatisme
 - La distance focale est différente selon l'axe X et l'axe Y car la lentille n'est pas parfaitement circulaire



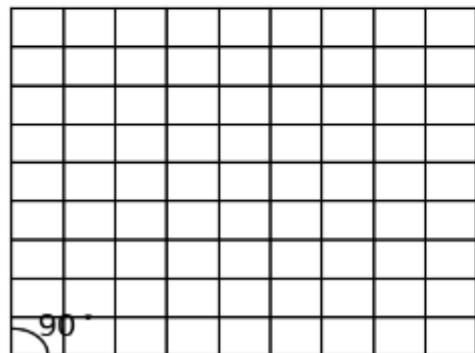
- Problème des lentilles
 - Aberration chromatique
 - Aberration comatique

Coma

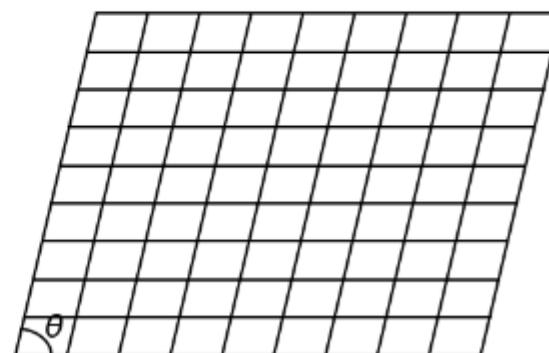


- Problème du capteur
 - Asymétrie des pixels ou « skewness »
 - Souvent négligeable sur les caméras modernes

Ideally aligned pixel grid (*Norm*)
Figure 1(a)



Skewed pixel grid (*Skew*)
Figure 1(b)

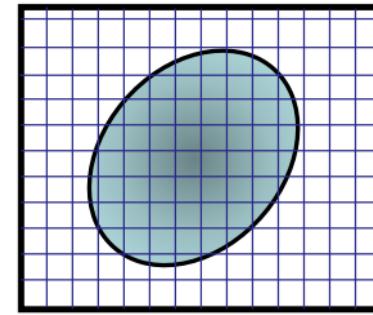
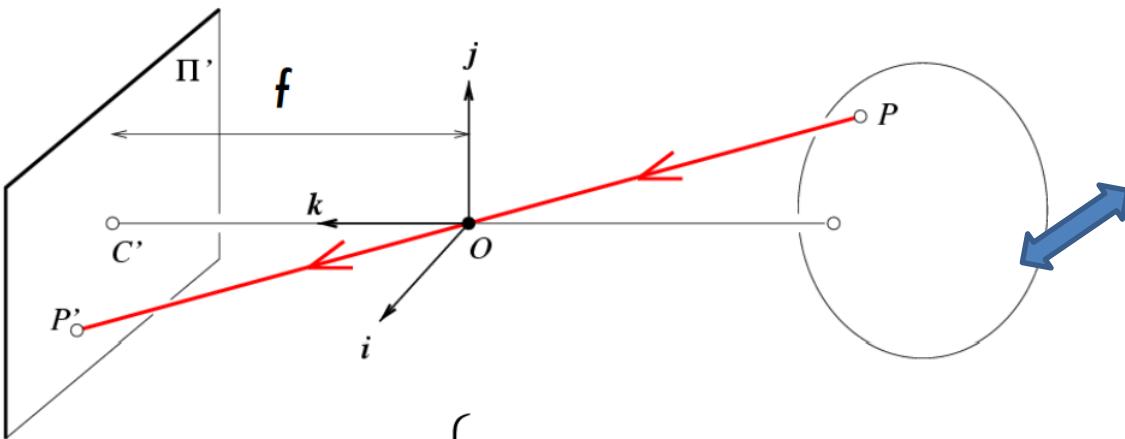


- Problème du capteur
 - Bruit lié à l'électronique
 - Bruit lié à la discrétisation des mesures (1 octet par pixel => seulement 255 valeurs possibles)



Propriété Magellium Atrial Group

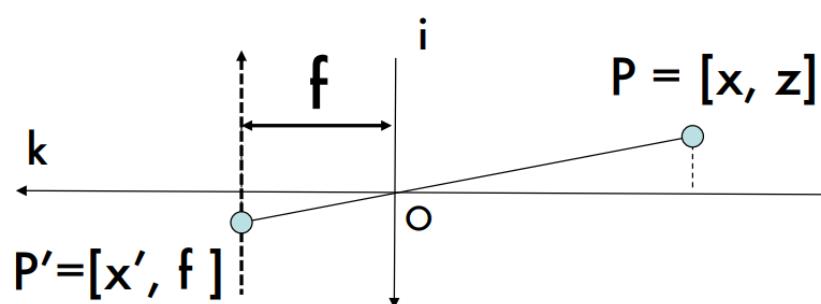




Digital image

Pixels, bottom-left coordinate systems

$$P = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \rightarrow P' = \begin{bmatrix} x' \\ y' \end{bmatrix} \quad \left\{ \begin{array}{l} x' = f \frac{x}{z} \\ y' = f \frac{y}{z} \end{array} \right. \quad \mathfrak{R}^3 \xrightarrow{E} \mathfrak{R}^2$$



f = focal length
o = center of the camera

$$\frac{x'}{f} = \frac{x}{z}$$

- 3D vers image 2D:

- $x' = f \cdot \frac{X}{Z}$

- $y' = f \cdot \frac{Y}{Z}$

- Passage du monde réel continu en mètres vers le monde discret de l'image en pixels

- $u = \frac{x'}{d_x} = \frac{f}{d_x} \cdot \frac{X}{Z}$

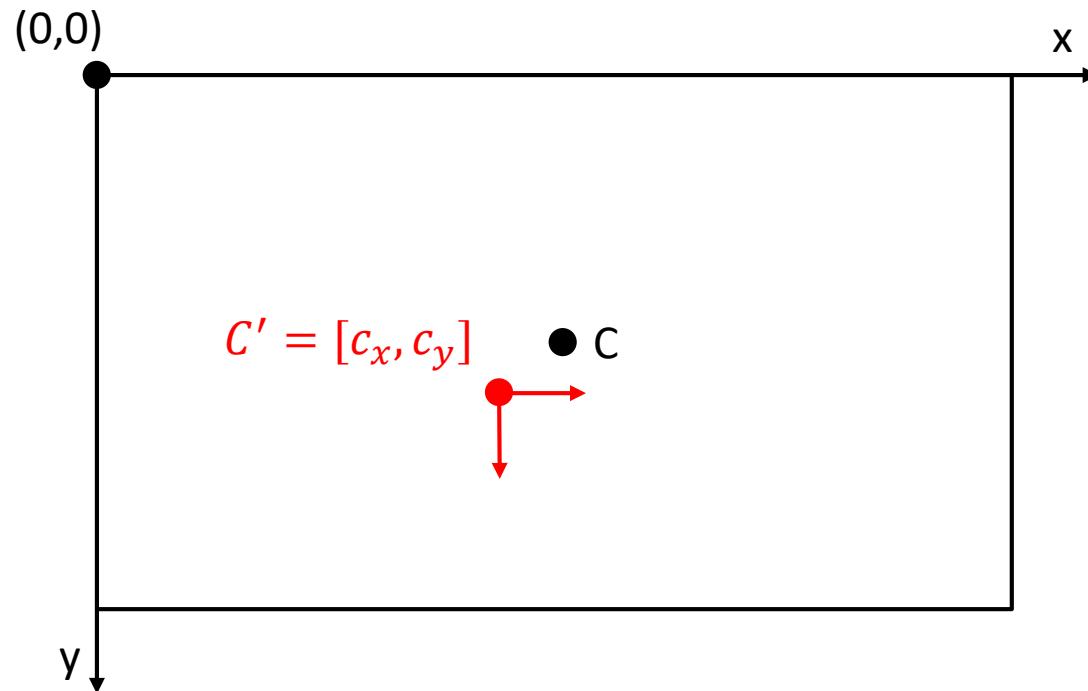
- $v = \frac{y'}{d_y} = \frac{f}{d_y} \cdot \frac{Y}{Z}$

- En réalité f est différent sur l'axe x et l'axe y

- $u = \frac{f_x}{d_x} \cdot \frac{X}{Z} = \alpha_x \cdot \frac{X}{Z}$

- $v = \frac{f_y}{d_y} \cdot \frac{Y}{Z} = \alpha_y \cdot \frac{Y}{Z}$

- Le capteur et la lentille ne sont pas parfaitement alignés => décalage entre le centre optique et le centre de l'image
- Le point/pixel de coordonnées (0,0) dans l'image corresponds au coin supérieur gauche
- => On applique une translation permettant de passer du centre optique au (0,0) de l'image



- 3D vers image 2D:

- $u = \alpha_x \cdot \frac{X}{Z} + c_x$
- $v = \alpha_y \cdot \frac{Y}{Z} + c_y$

- Matrice intrinsèque:

- $K = \begin{bmatrix} \alpha_x & 0 & c_x \\ 0 & \alpha_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$

- Souvent on trouvera la notation:

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

, s le coefficient skew, souvent négligeable

Matrice intrinsèque

avec ici $f_x = \alpha_x$ et $f_y = \alpha_y$ les focales en pixels

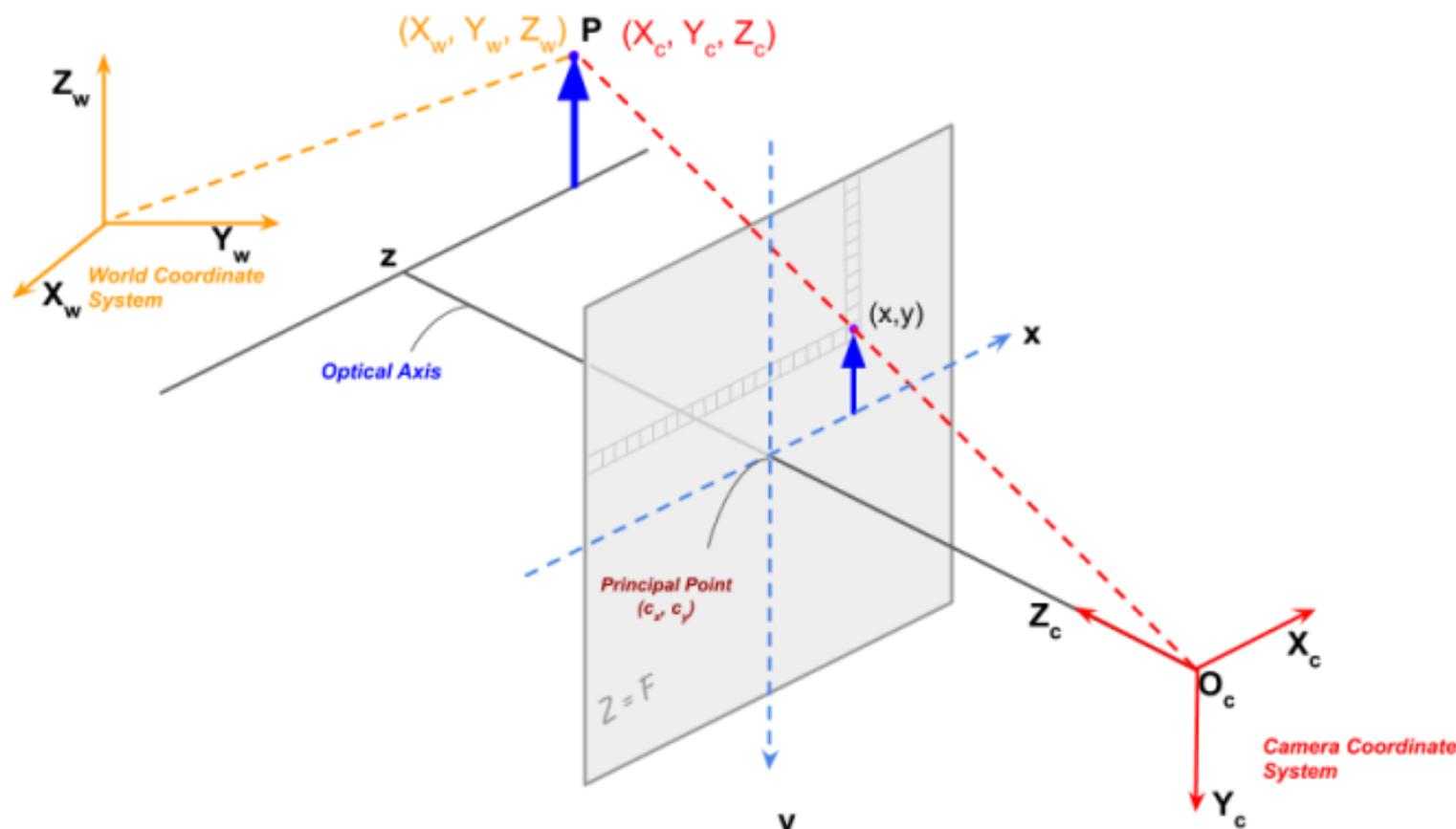
- 3D vers image 2D:

- $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim K \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \Rightarrow$ Équivalent à un facteur d'échelle près
- $\exists \lambda \text{ tq. } \lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$

- Preuve:

- $K \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \alpha_x & 0 & c_x \\ 0 & \alpha_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \alpha_x X + c_x Z \\ \alpha_y Y + c_y Z \\ Z \end{bmatrix}$
- $\begin{bmatrix} \alpha_x X + c_x Z \\ \alpha_y Y + c_y Z \\ Z \end{bmatrix} \sim \begin{bmatrix} \alpha_x \frac{X}{Z} + c_x \\ \alpha_y \frac{Y}{Z} + c_y \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$

- Les coordonnées du monde ne sont pas nécessairement les mêmes que les coordonnées de la caméra



- Il faut convertir du systèmes de coordonnées du monde vers le système de coordonnées de la caméra:

- $$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t$$
 - R une matrice de rotation 3x3
 - t un vecteur translation

- Notation matricielle:

- $$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = [R|t] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Rightarrow \bar{X}_c = [R|t] \cdot \bar{X}$$
 - $[R|t]$ matrice 3x4
- On peut aussi noter $\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$ en coordonnées homogènes

- On a:

- Les paramètres intrinsèques: $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim_K \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}$

- Les paramètres extrinsèques: $\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = [R|t] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$

- Donc:

- $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim_K [R|t] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = P \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$

- Avec P la matrice caméra 3×4

- D'où: $\bar{x} \sim P \bar{X}$

- P une matrice $3 \times 4 \Rightarrow 12$ inconnues

- $P \sim K \cdot [R|t]$,
- on peut décomposer P en $P = [P_1|P_2]$
- $P_1 = K \cdot R$, matrice 3×3
- $P_2 = K \cdot t$, vecteur 3×1

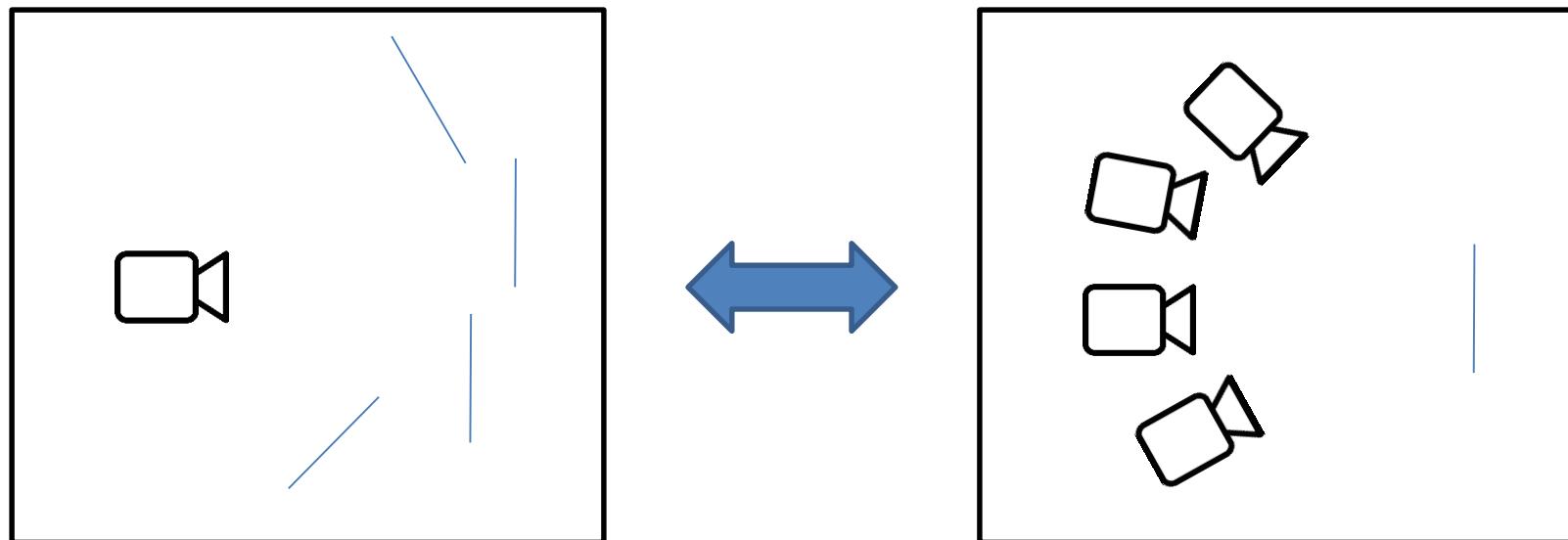
- Resectioning:

- On estime P à partir de paires \bar{x} et \bar{X} connues
- $\bar{x} \sim P\bar{X}$ donne deux équations indépendantes à 12 inconnues pour chaque point
- On résous $Ap = 0$
 - Avec p le vecteur de 12 inconnues
 - A la matrice $2n \times 12$ représentant les équations indépendantes
 - n le nombre de points connus
- Il faut connaître 6 points distincts minimum

Estimation de la matrice intrinsèque

- Plane-based calibration:

- On réalise l'acquisition de multiple images d'une surface plane (e.g. damier)

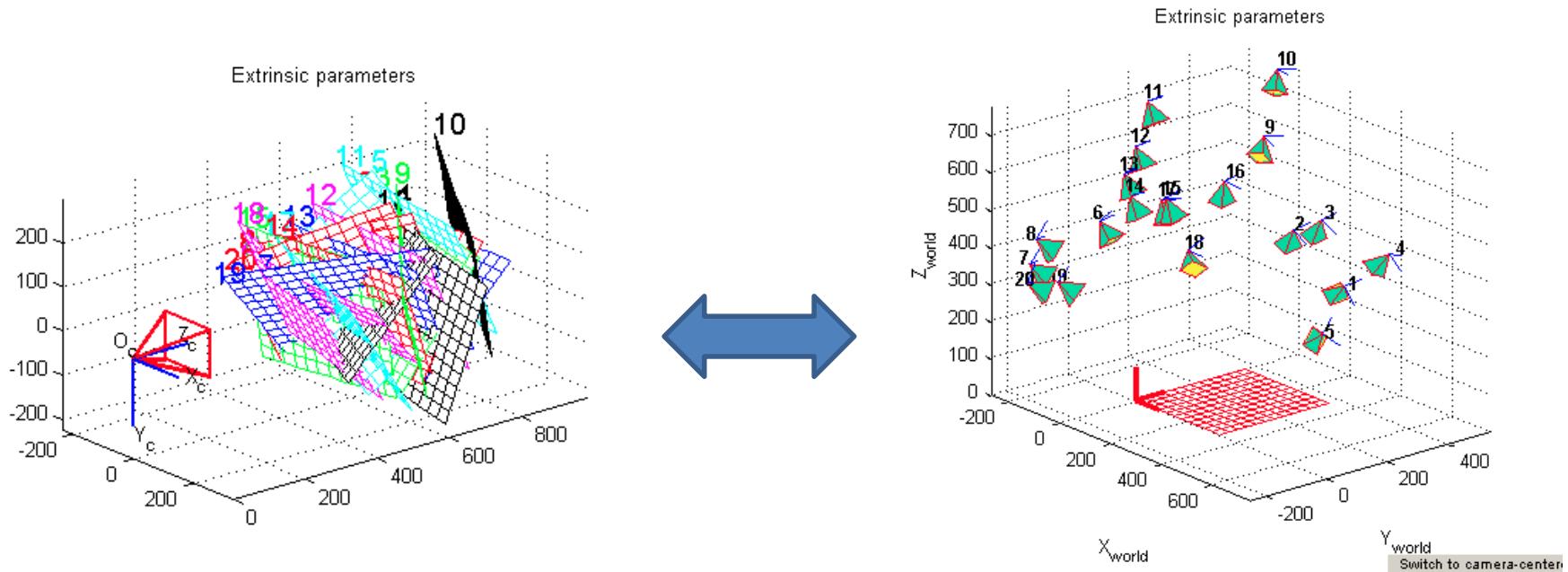


- Cela permet de fixer $Z=0$ pour tous les points du plan observé

Estimation de la matrice intrinsèque

- Plane-based calibration:

- On réalise l'acquisition de multiple images d'une surface plane (e.g. damier)

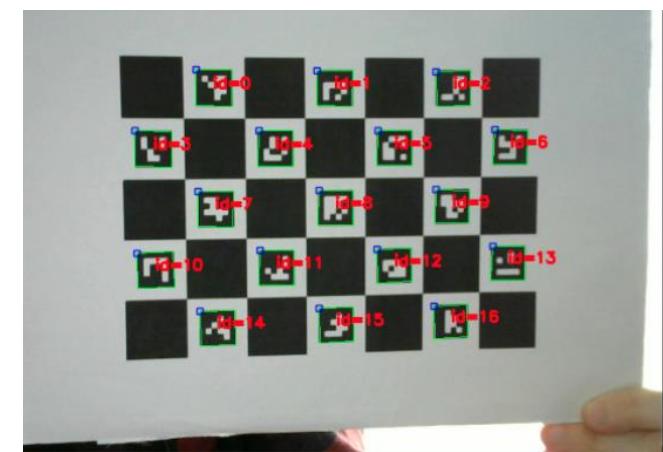
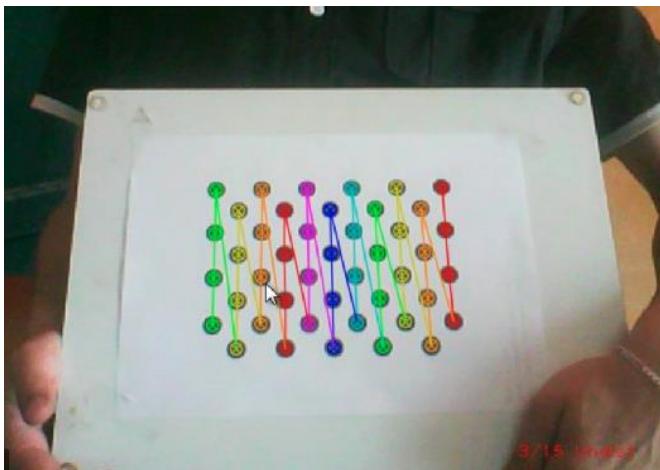


- Cela permet de fixer $Z=0$ pour tous les points du plan observé

Estimation de la matrice intrinsèque

- On a donc:

- $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim K \cdot [R|t] \cdot \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim K[r_1 \quad r_2 \quad t] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$
- À l'aide de ces équations et de la connaissance du plan observé, il est possible de déterminer K.



- Il est aussi possible d'estimer des paramètres caractérisant les distorsions radiales et tangentielles.
- Plusieurs modèles de distorsion existent, notamment le modèle polynomial de Brown-Conrady:

- Distorsions radiales:

- $x_{distorted} = x \cdot \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6}$
 - $y_{distorted} = y \cdot \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6}$
 - Avec $r = \sqrt{x^2 + y^2}$ et $x = \frac{X_c}{Z_c}$, $y = \frac{Y_c}{Z_c}$

- Distorsions tangentielles:

- $x_{distorted} = x + 2p_1xy + p_2(r^2 + 2x^2)$
 - $y_{distorted} = y + p_1(r^2 + 2y^2) + 2p_2xy$

- Modèle de distorsions de Brown-Conrady:

- $x_{distorted} = x \cdot \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6} + 2p_1xy + p_2(r^2 + 2x^2)$
- $y_{distorted} = y \cdot \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6} + p_1(r^2 + 2y^2) + 2p_2xy$

- Rappel, sans distorsions on a:

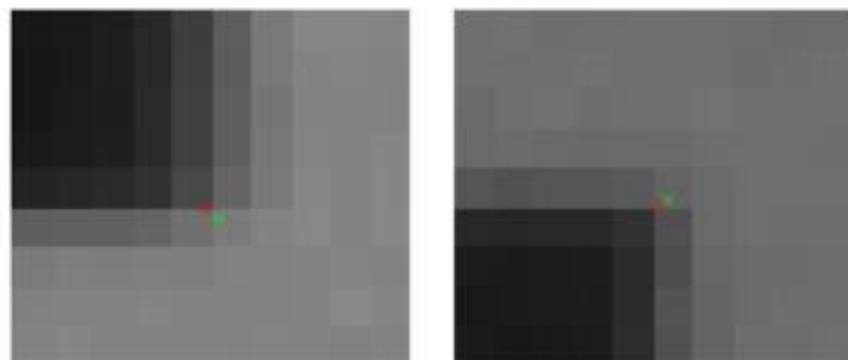
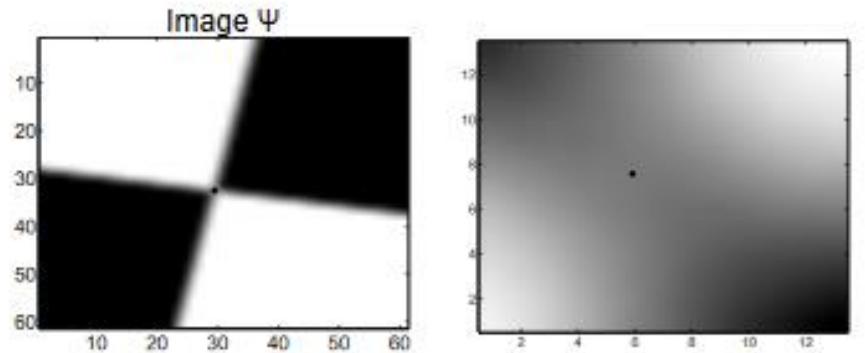
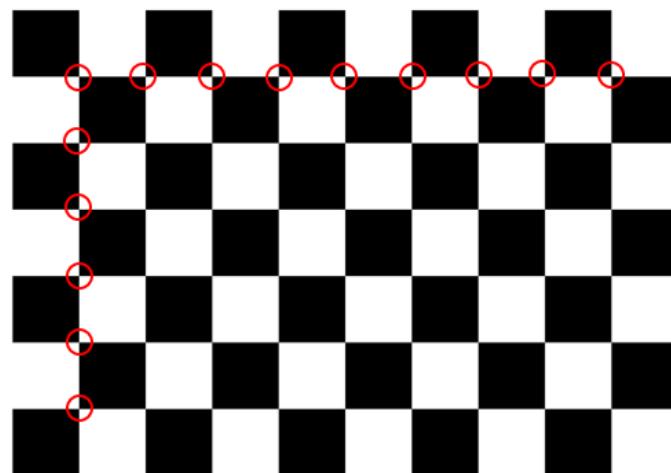
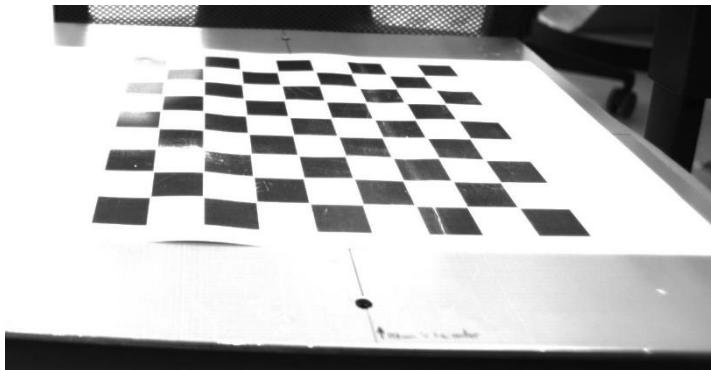
- $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x \cdot x + c_x \\ \alpha_y \cdot y + c_y \\ 1 \end{bmatrix} \sim K \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = K \cdot [R|t] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = P \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Rightarrow \bar{x} \sim P\bar{X}$
- avec $x = \frac{X_c}{Z_c}$, $y = \frac{Y_c}{Z_c}$

- Modèle complet:

- $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x \cdot x_{distorted} + c_x \\ \alpha_y \cdot y_{distorted} + c_y \\ 1 \end{bmatrix}$, équations non linéaires!

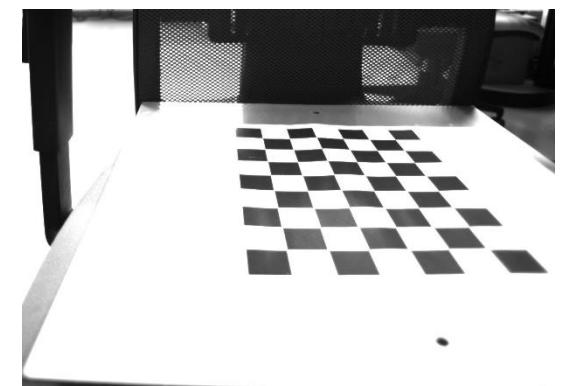
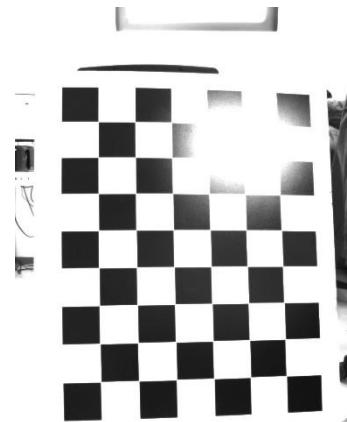
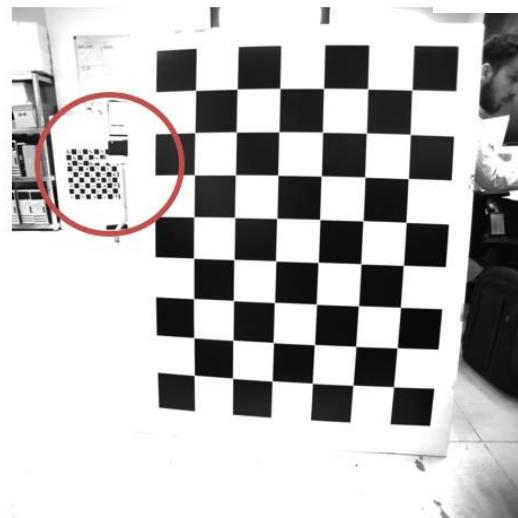
Comment réaliser une bonne calibration?

- La cible doit être parfaitement plane
- Les motifs de type cercles asymétriques donnent de meilleurs résultats
- La distance entre les points doit être mesurée précisément
- Il faut correctement définir le nombre de lignes/colonnes de la cible



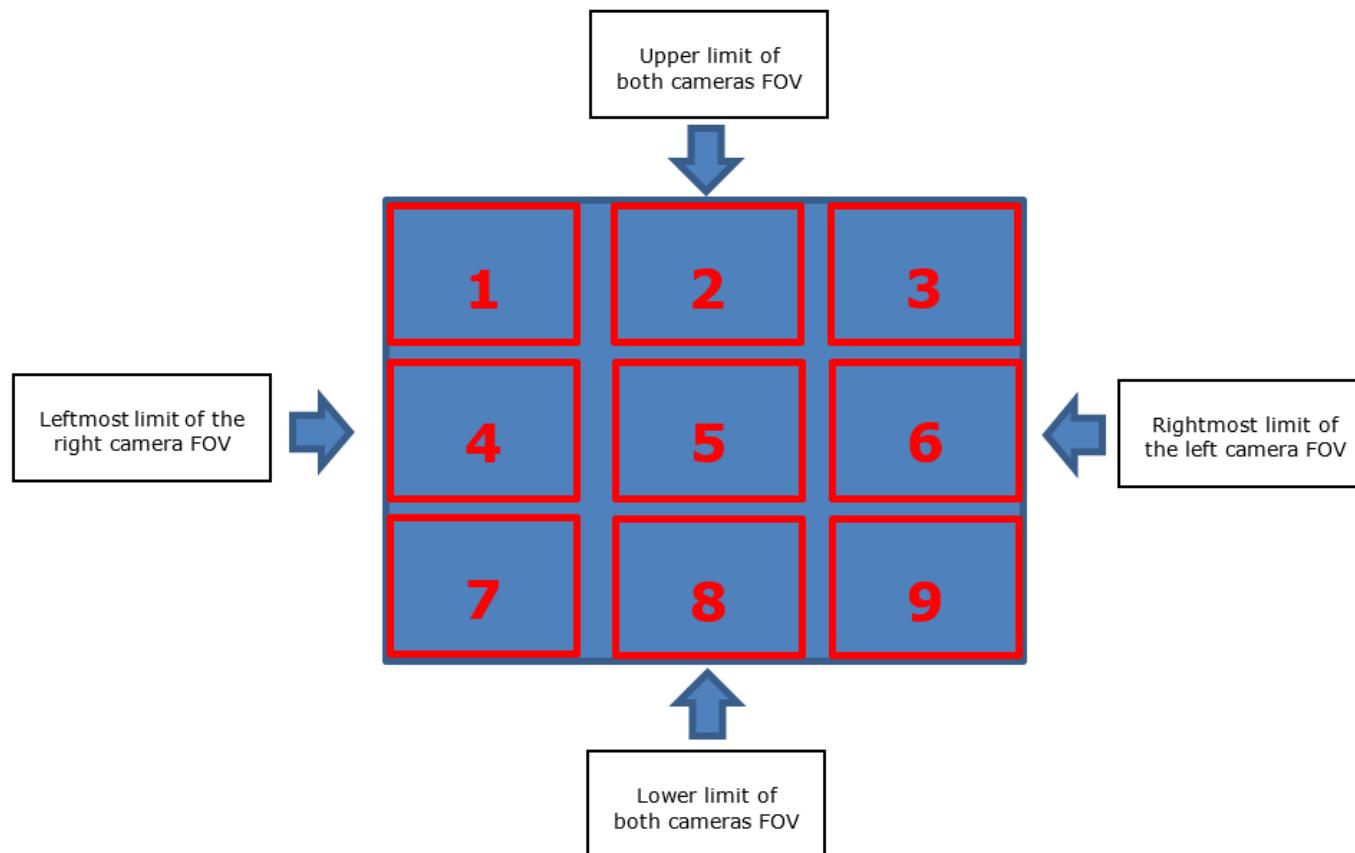
Contrôler l'environnement de calibration

- Pas de sources de lumière directe
- Pas d'autre cible/motifs similaires visible
- Pas de reflets
- Contrôle de l'exposition



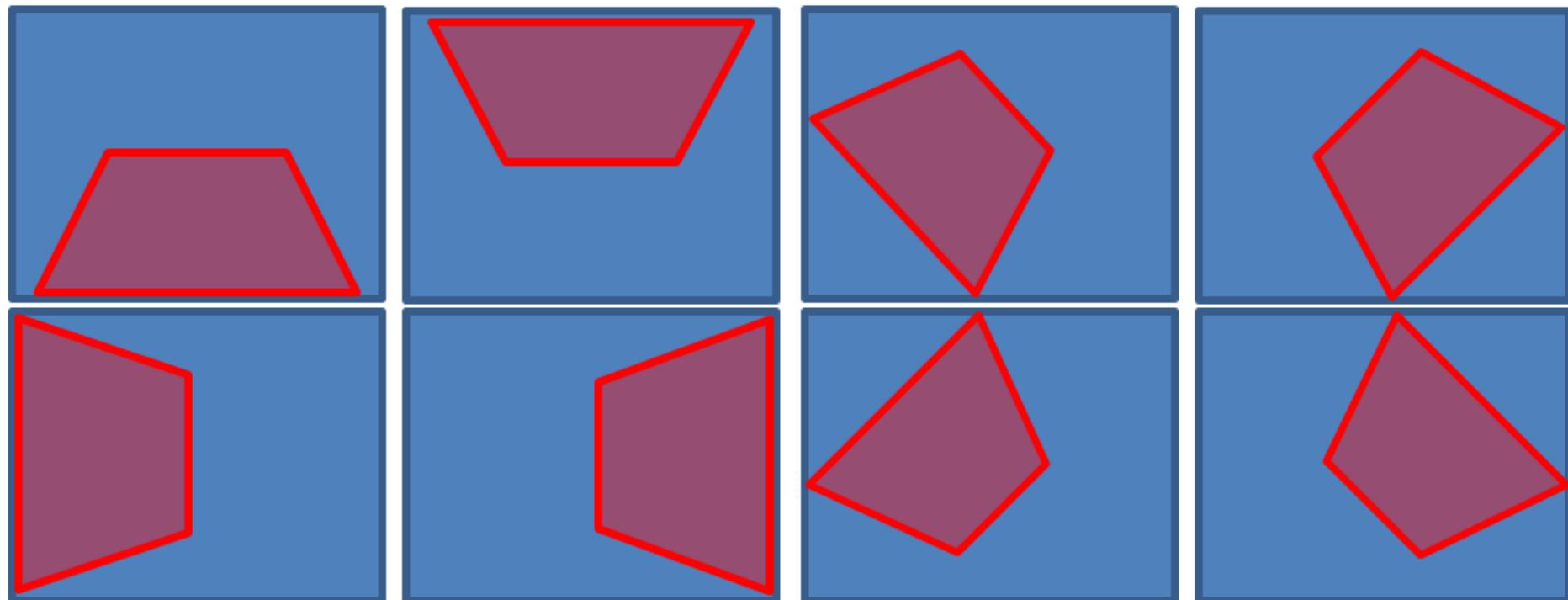
Procédure de calibration

- Mettre la caméra dans une position fixe
- Acquérir 9 images à la distance de travail la plus proche
- Recommencer pour la distance de travail moyenne
- Recommencer pour la distance de travail la plus lointaine



Procédure de calibration

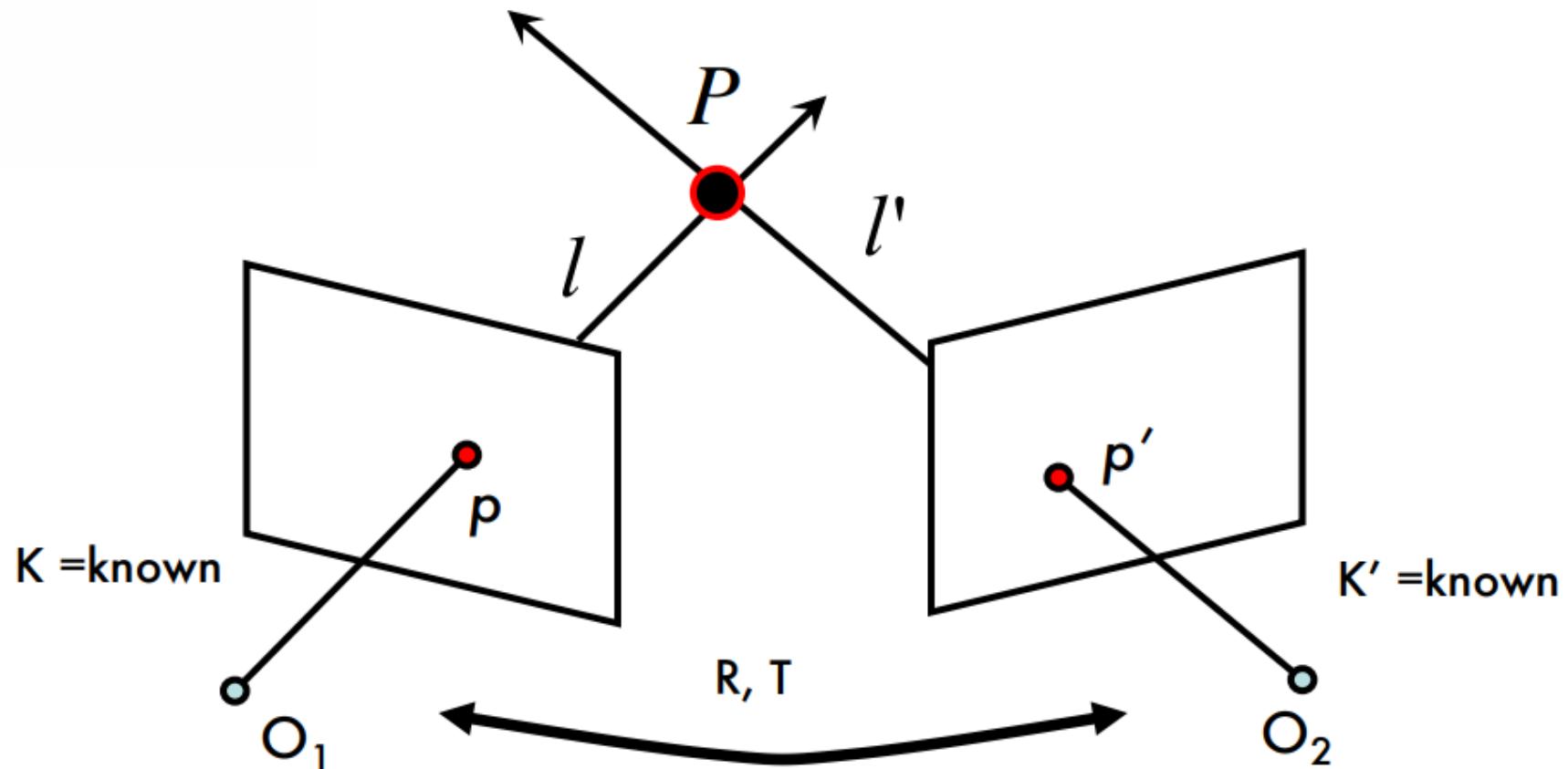
- Acquérir 8 images avec des inclinaisons différentes de la cible
- Acquérir 5 à 10 images supplémentaire avec des angles « aléatoires »
- Essayer d'avoir une distribution homogène des points dans le plan image



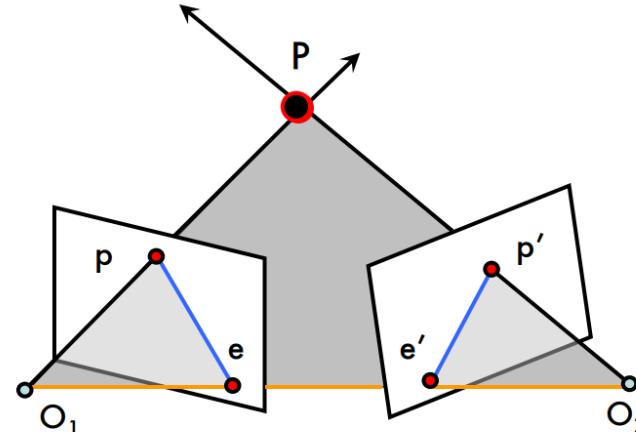
Partie I: Perception 3D

- I. Introduction
- II. Systèmes d'acquisition 3D
- III. Modélisation de la camera
- IV. Stéréovision

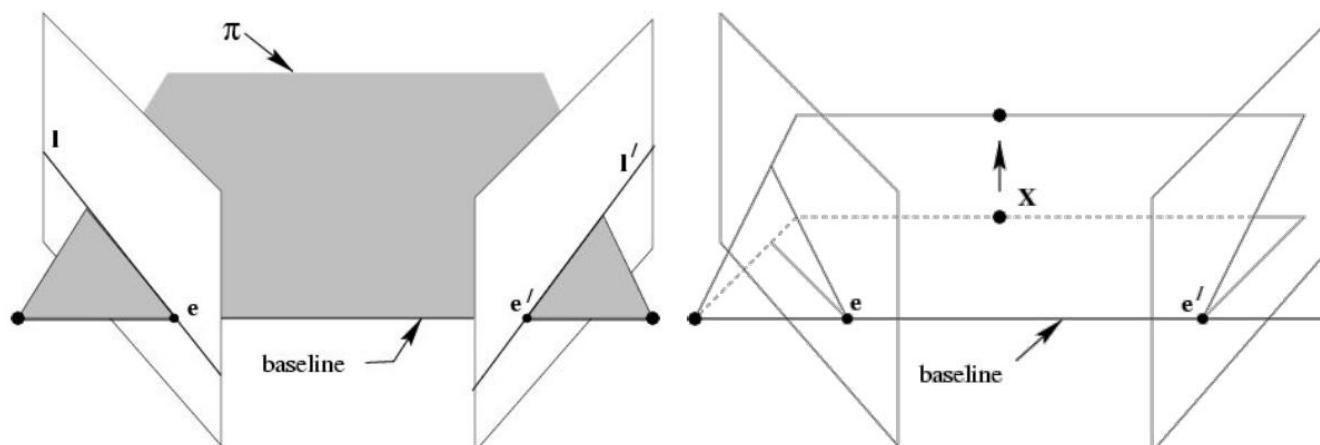
Triangulation



Géométrie épipolaire



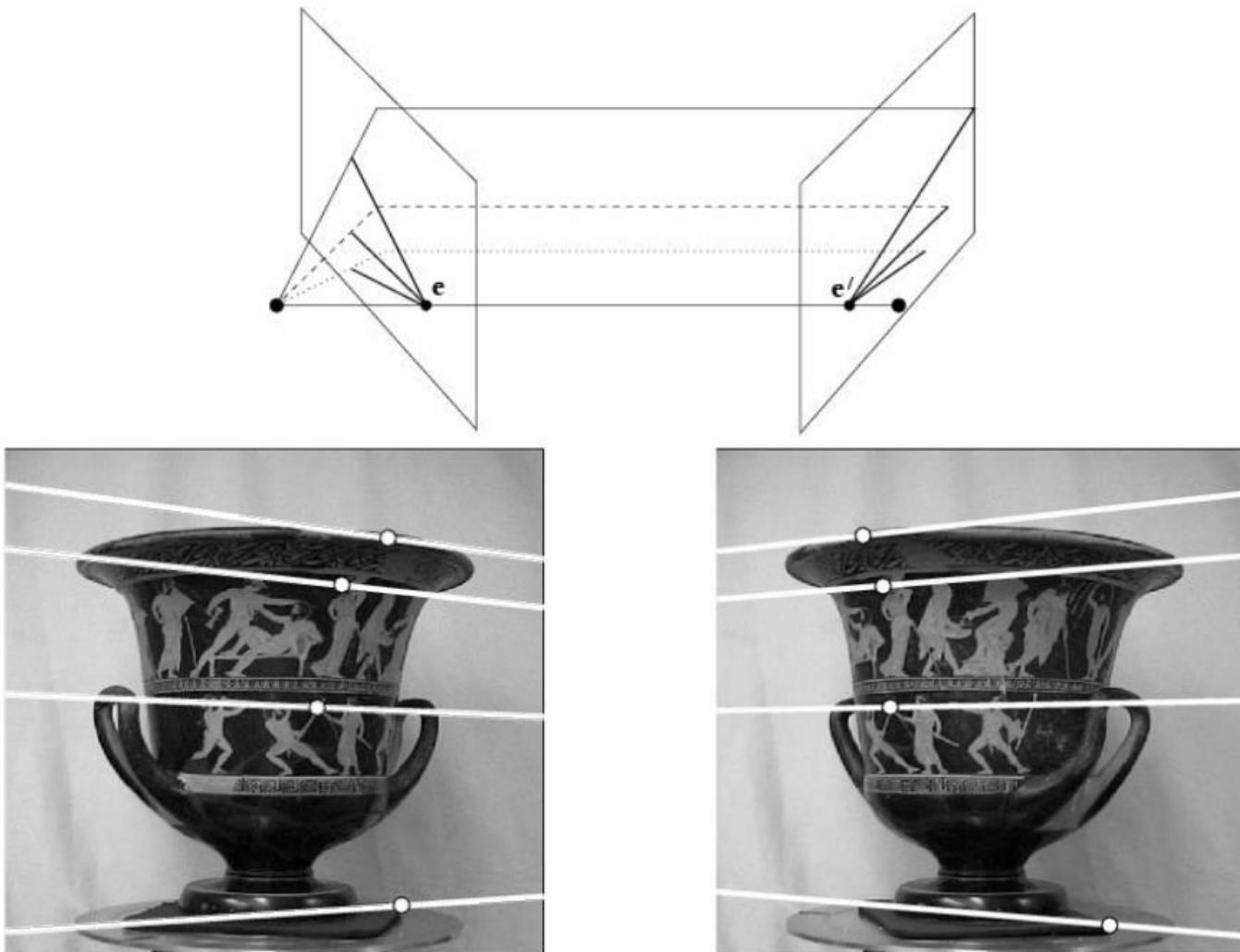
- Epipolar Plane
- Baseline
- Epipolar Lines
- Epipoles e, e'
 - = intersections of baseline with image planes
 - = projections of the other camera center



Tous les points sur π se projettent sur l et l'

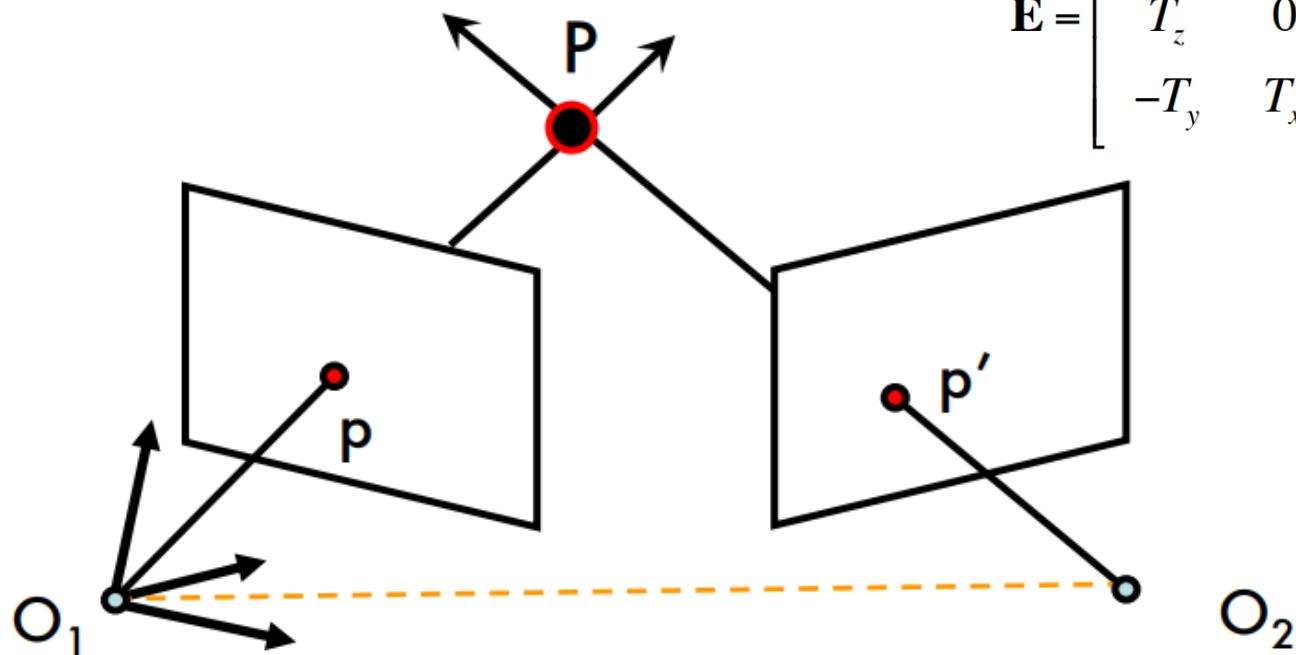
Famille de plans π et lignes l et l'
Se coupent en e et e'

Lignes épipolaire



Matrice essentielle E

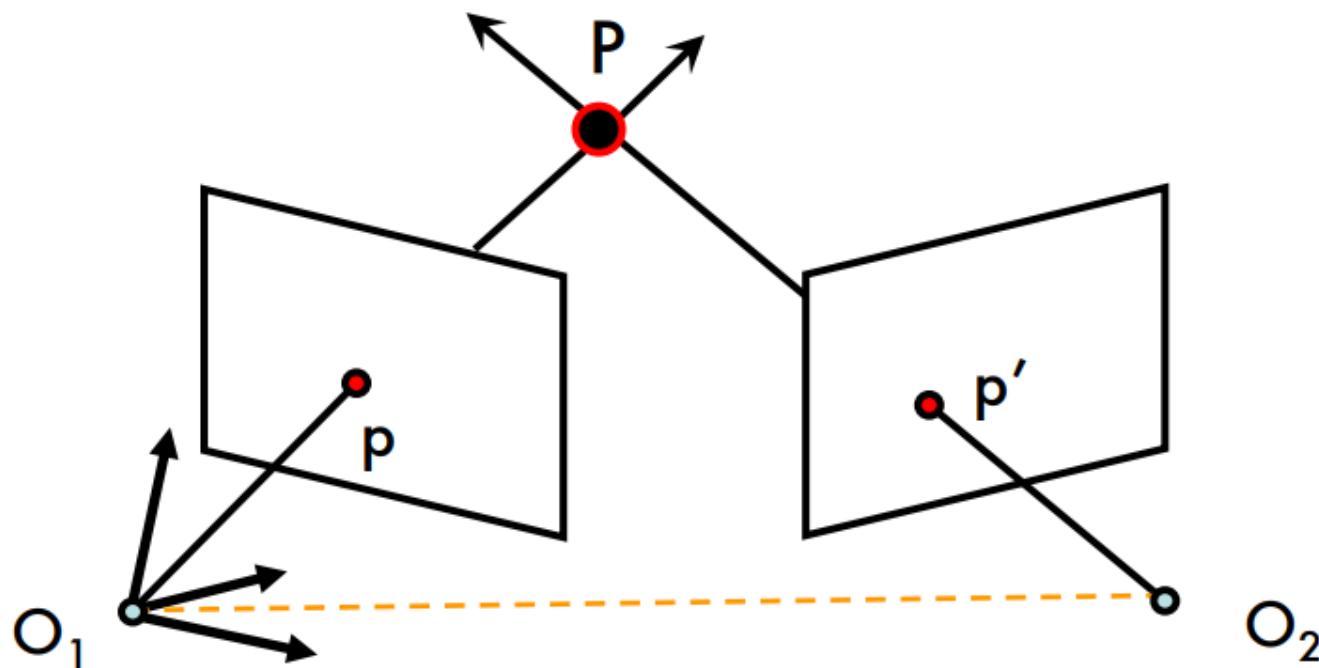
$$\mathbf{E} = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \mathbf{R}$$



$$p^T E p' = 0$$

$$E = [T_x] \cdot R$$

E = Essential Matrix

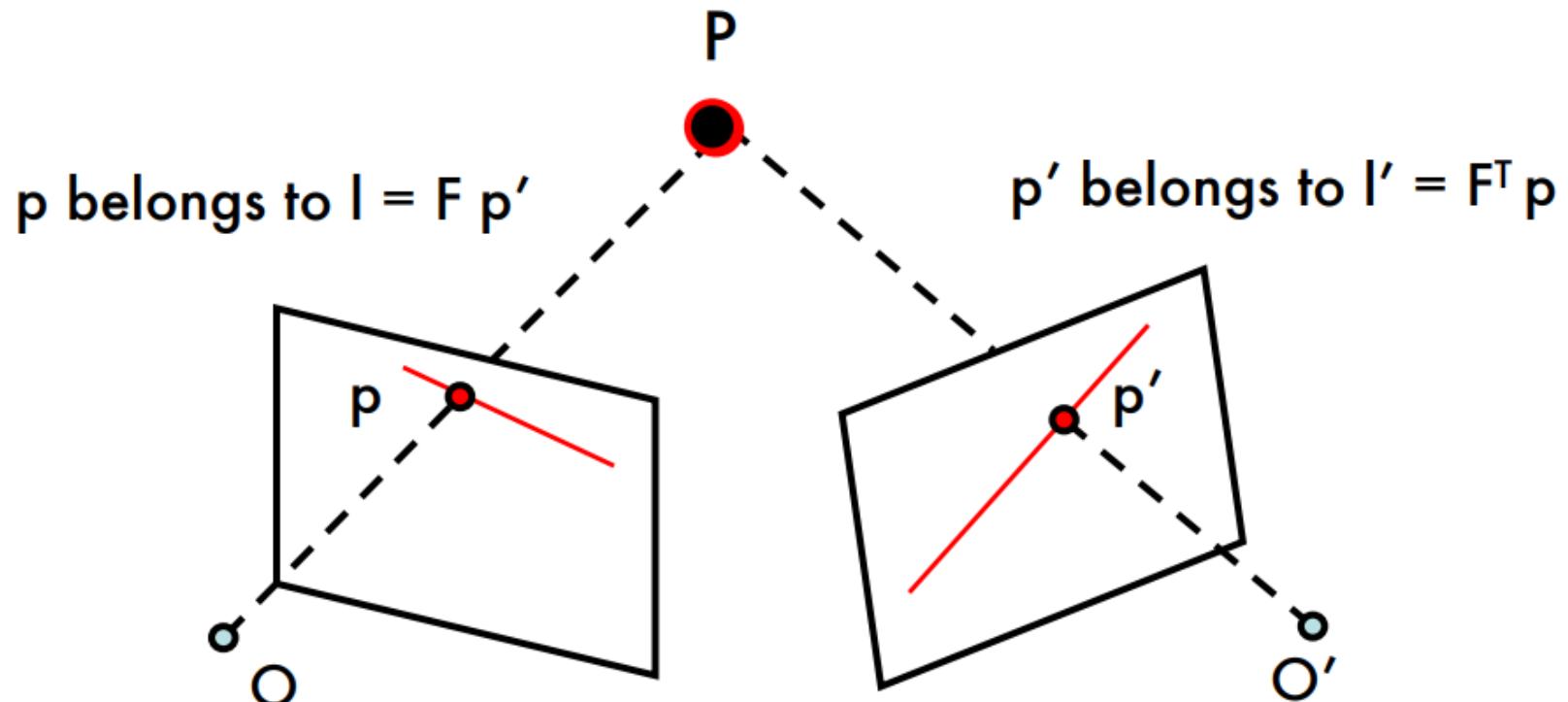


$$p^T F p' = 0$$

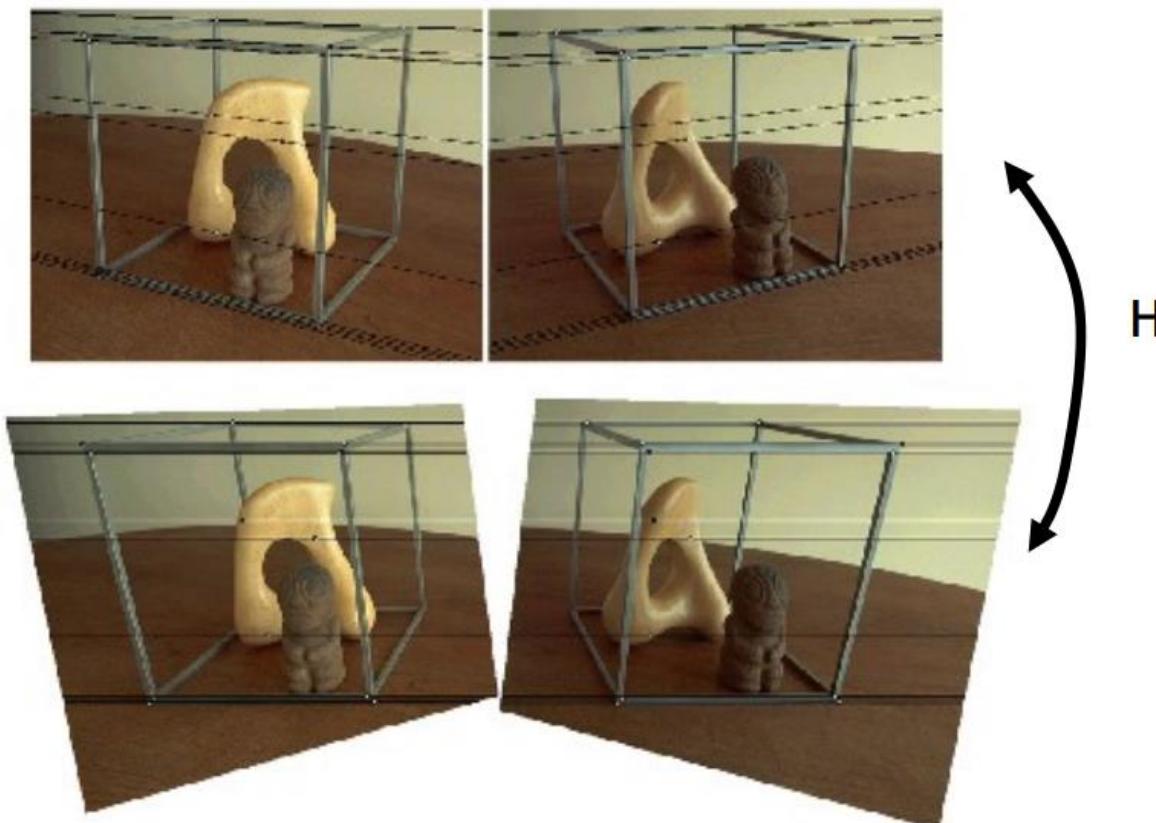
$$F = K^{-T} \cdot [T_x] \cdot R \cdot K'^{-1}$$

F = Fundamental Matrix

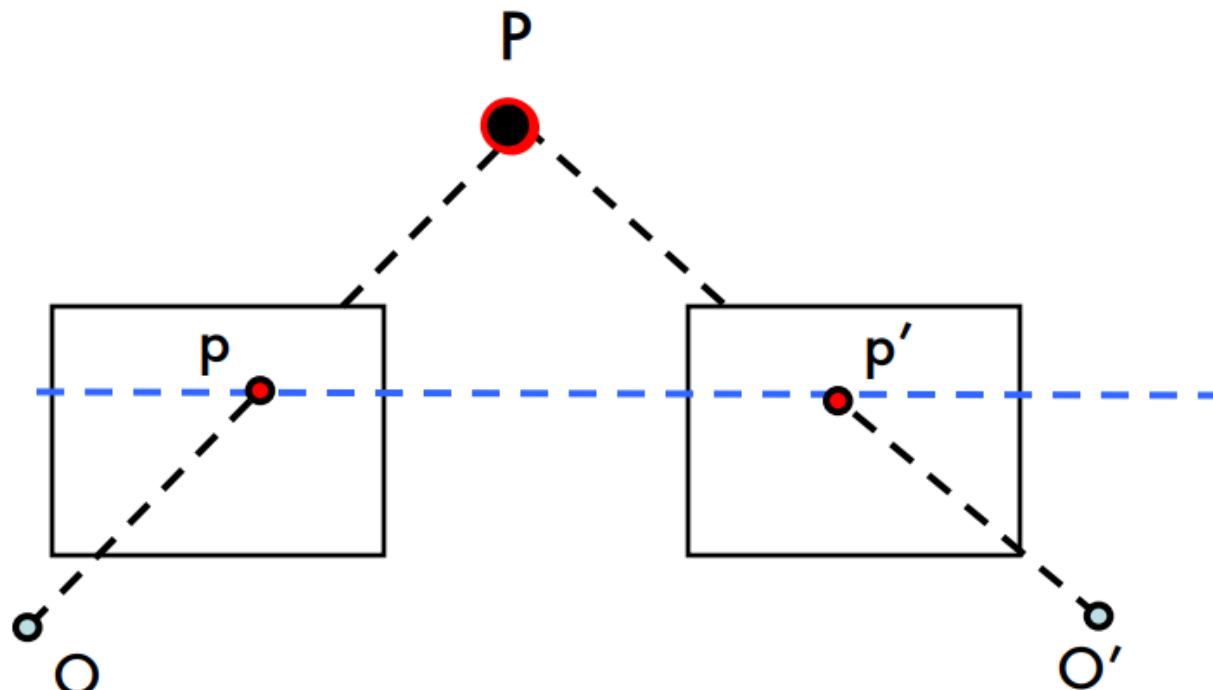
- Trouver les coordonnées en pixel d'un point 3D dans le plan image des deux caméras



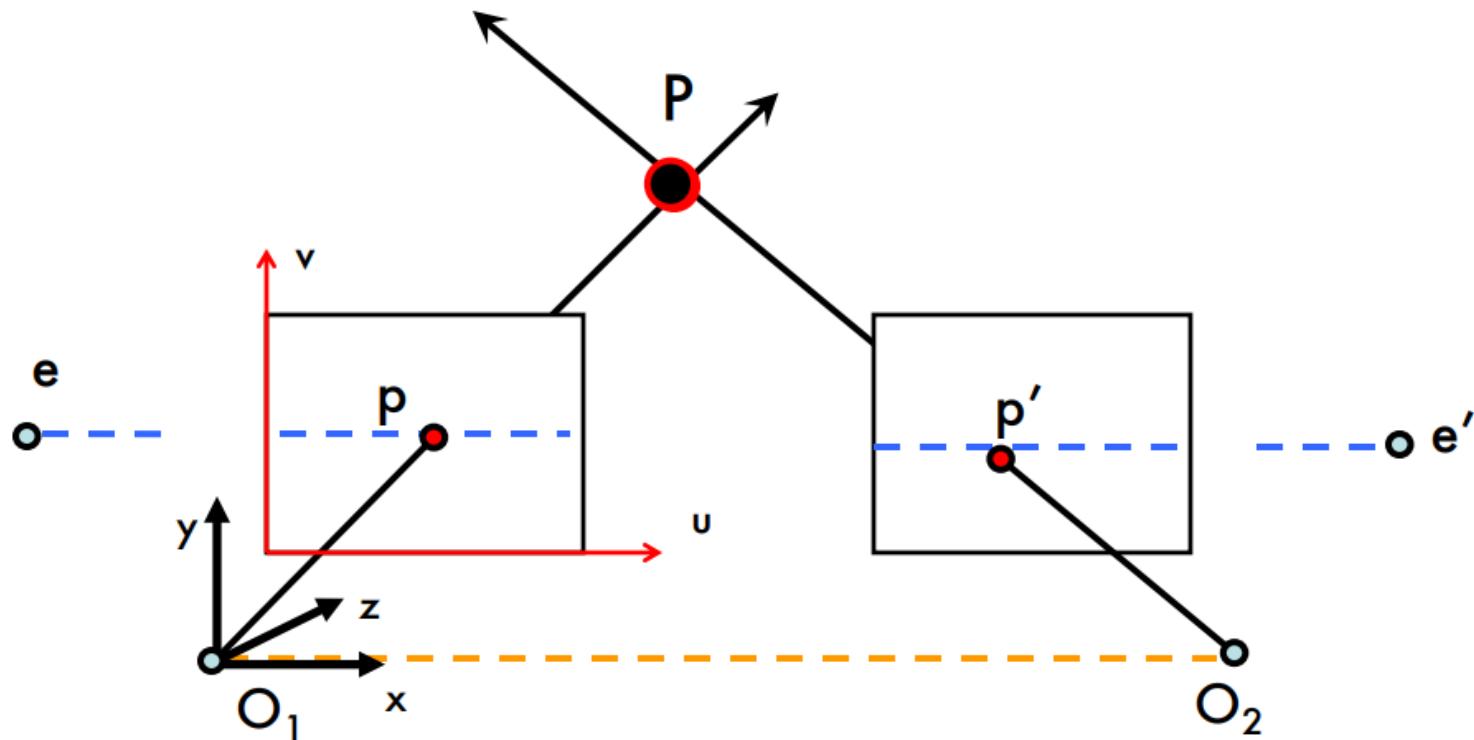
- Le fait de rendre deux images « parallèles »
 - Rends la triangulation facile
 - Facilite l'appariement



- Trouver les coordonnées en pixel d'un point 3D dans le plan image des deux caméras
- Les points images sont sur la même ligne lorsque les images sont rectifiées



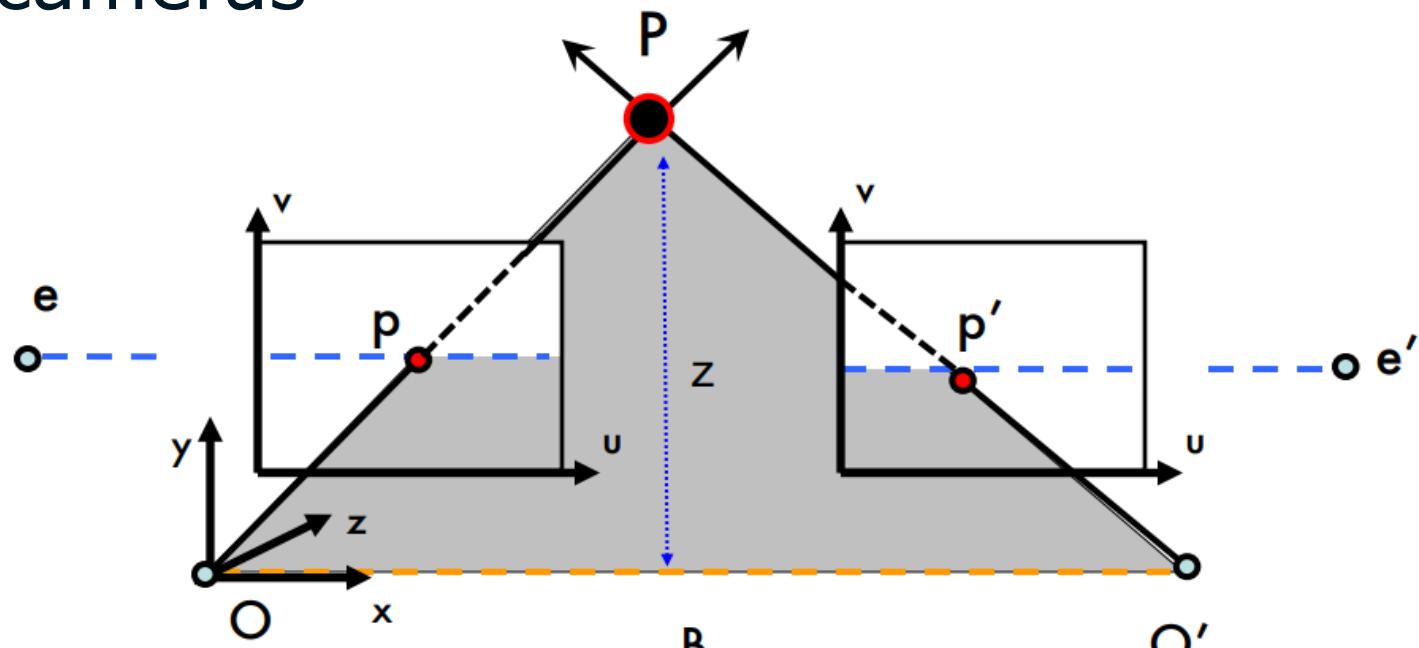
Plans image parallèles



- Epipolar lines are horizontal
- Epipoles go to infinity
- v -coordinates are equal

$$p = \begin{bmatrix} p_u \\ p_v \\ 1 \end{bmatrix} \quad p' = \begin{bmatrix} p'_u \\ p'_v \\ 1 \end{bmatrix}$$

- Disparité = distance en pixels qui sépare la projection d'un même point sur les images des deux caméras

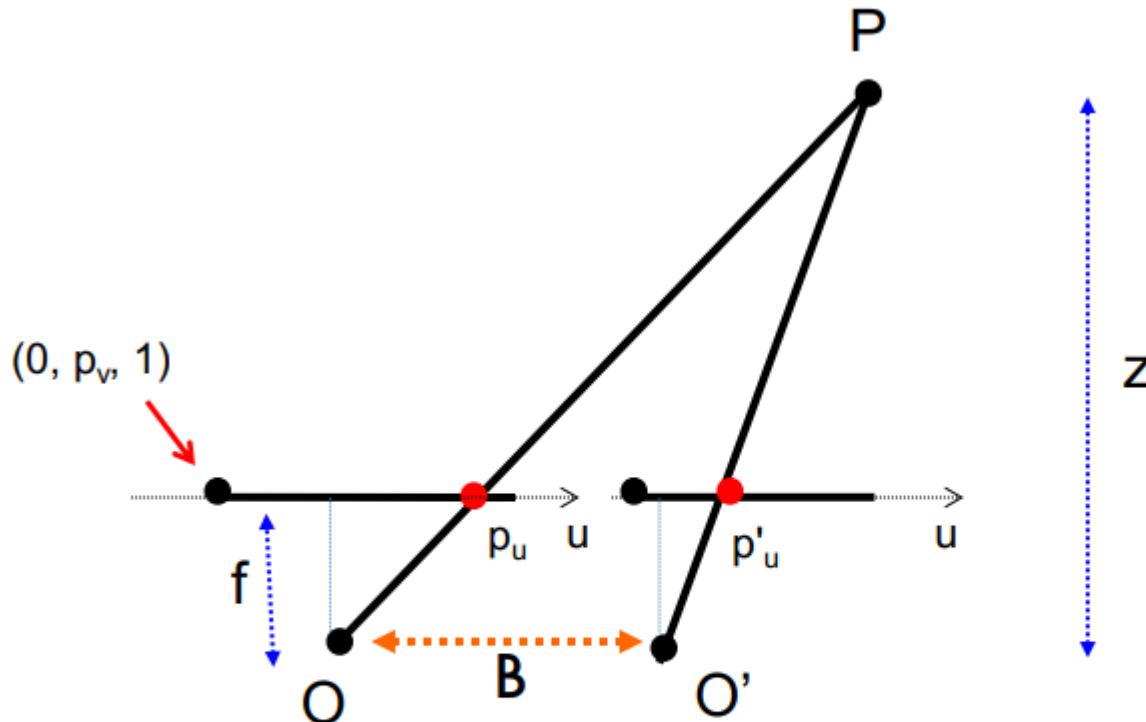


$$p = \begin{bmatrix} p_u \\ p_v \\ 1 \end{bmatrix} \quad p' = \begin{bmatrix} p'_u \\ p'_v \\ 1 \end{bmatrix}$$

$$\text{disparity} = p_u - p'_u \propto \frac{B \cdot f}{z}$$

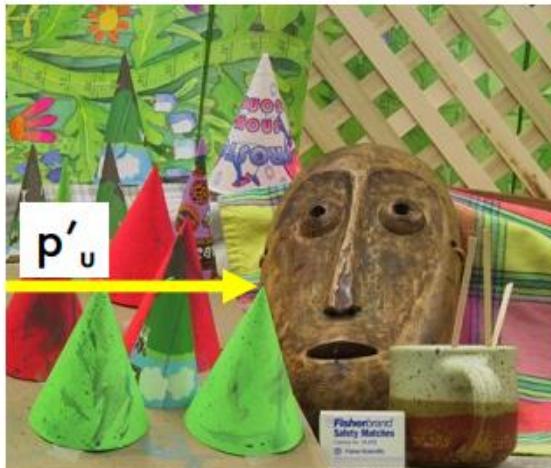
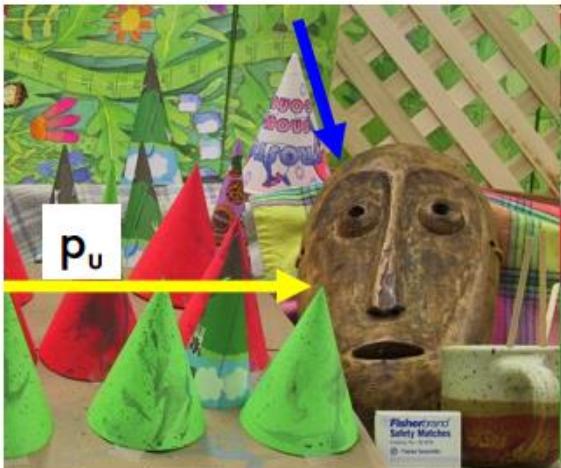
Disparity is inversely proportional to depth z !

Calcul de la profondeur



$$\text{disparity} = p_u - p'_u \propto \frac{B \cdot f}{z} \rightarrow z = \frac{B f_x}{d}$$

Image de disparité/carte de profondeur

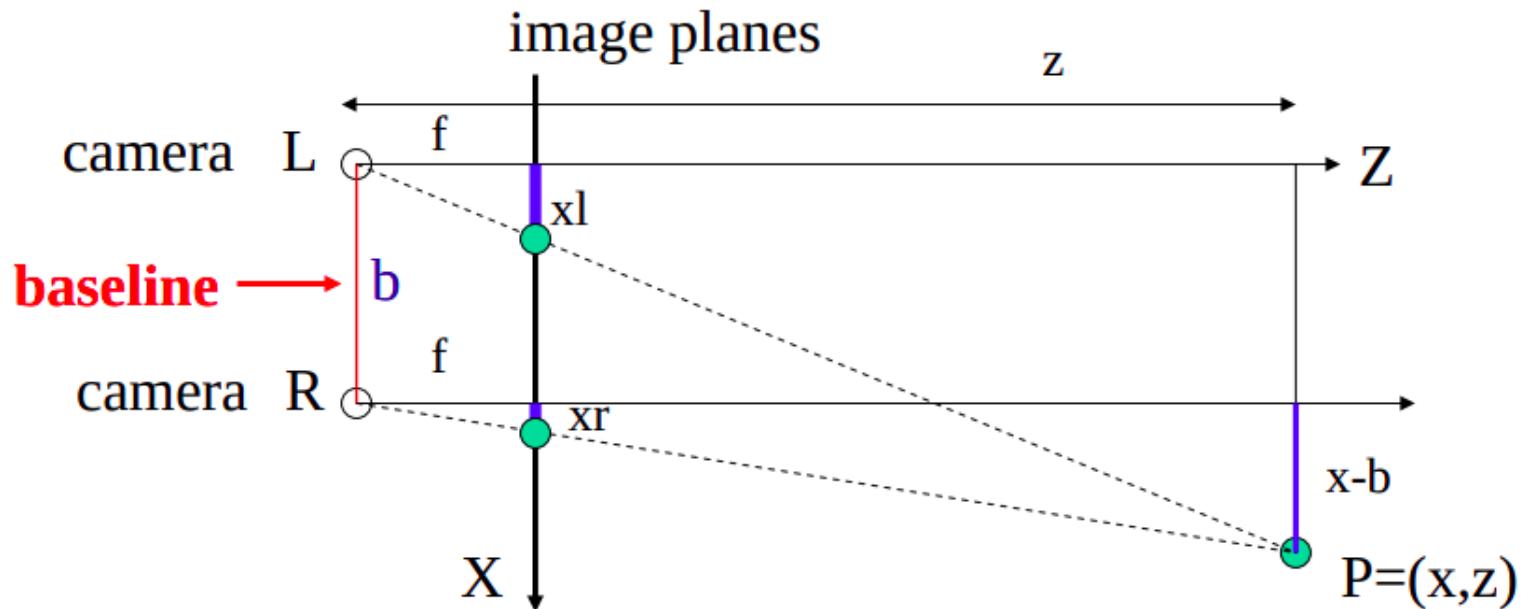


$$p_u - p'_u \propto \frac{B \cdot f}{z}$$

Stereo pair



Disparity map / depth map



- Théorème de Thalès:

- $\frac{d}{B} = \frac{f_x}{Z}$
- $\frac{x_L}{X} = \frac{x_R}{X-B} = \frac{f_x}{Z}$
- $\frac{y_L}{Y} = \frac{y_R}{Y} = \frac{f_y}{Z}$

$$\begin{aligned} &\Rightarrow Z = \frac{B}{d} \cdot f_x \\ &\Rightarrow X = x_L \cdot \frac{B}{d} = (x_R + d) \cdot \frac{B}{d} \\ &\Rightarrow Y = y_L \cdot \frac{B}{d} \cdot \frac{f_x}{f_y} = y_R \cdot \frac{B}{d} \cdot \frac{f_x}{f_y} \end{aligned}$$

- Comment faire pour trouver les points correspondants?

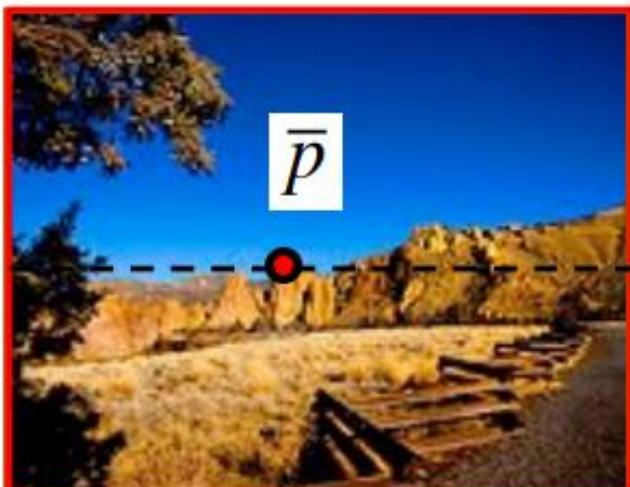


image 1

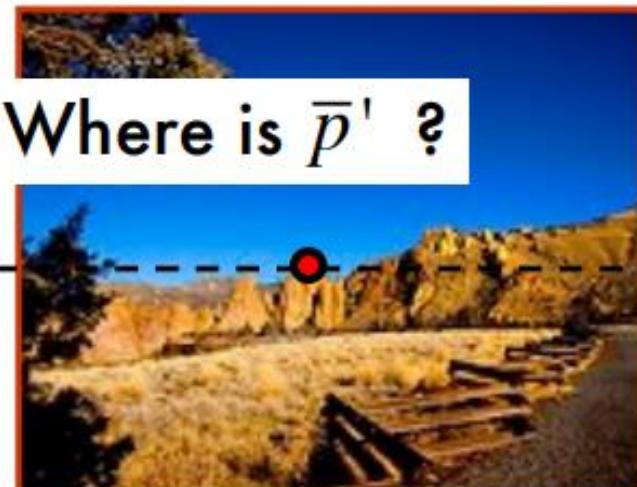
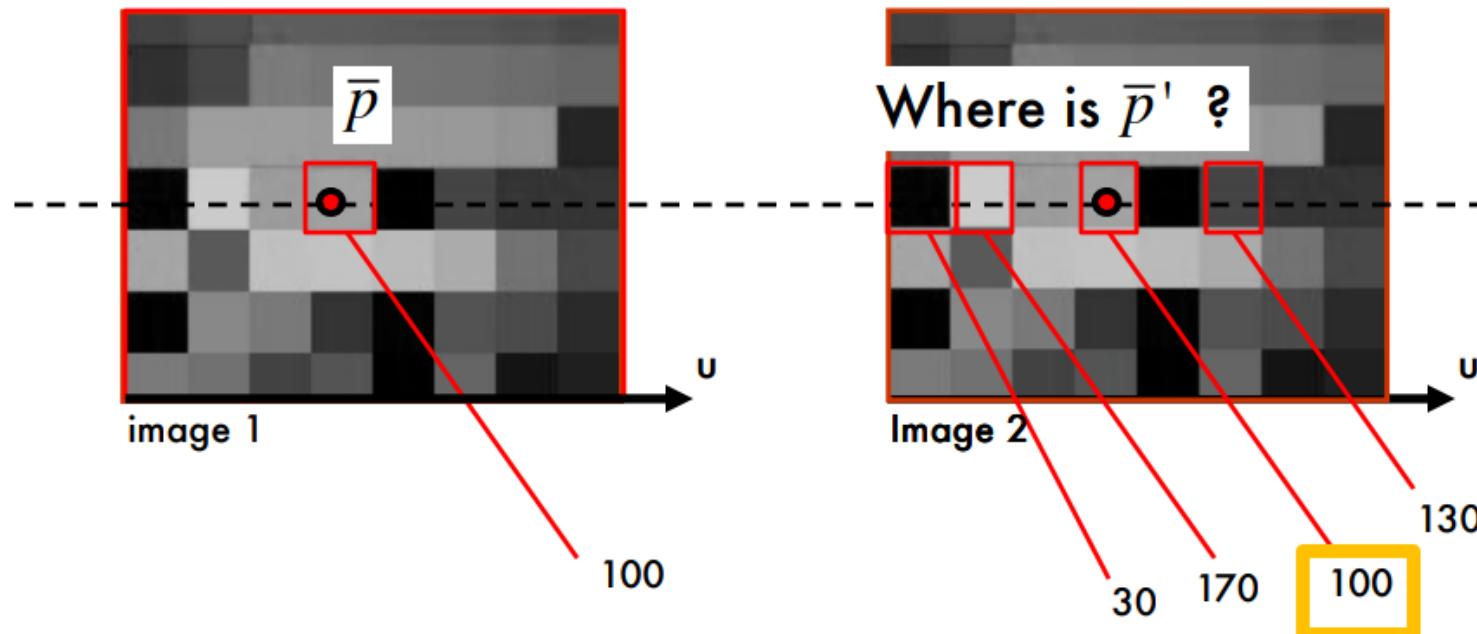
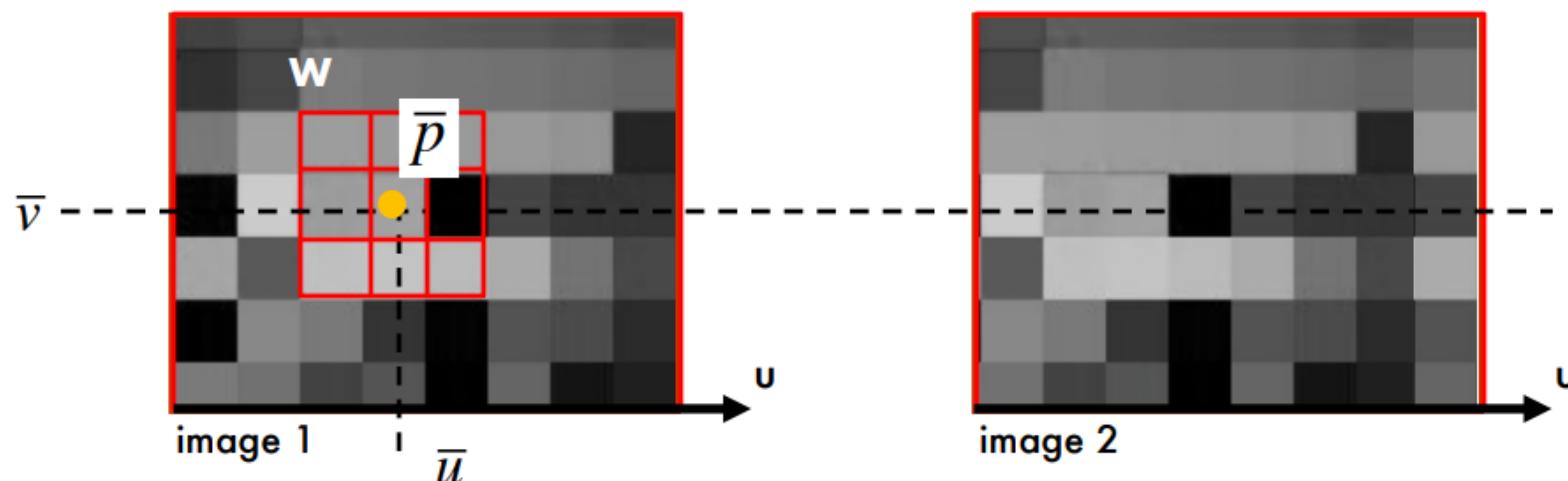


Image 2

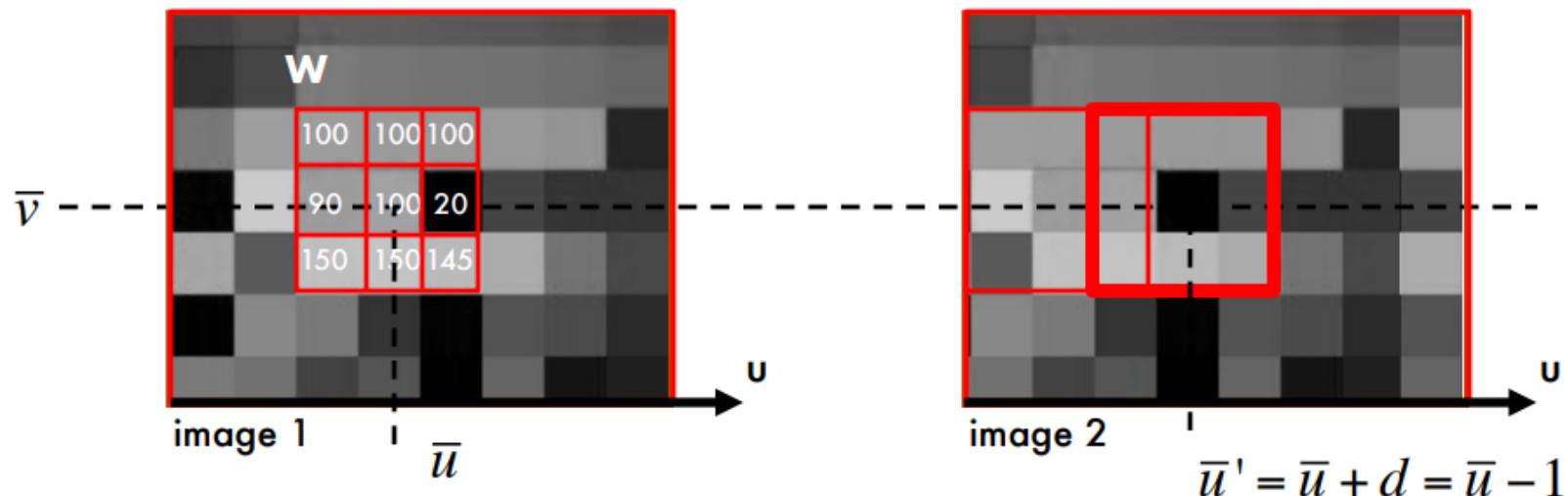
- Recherche de valeur identique



- Sélection d'une fenêtre de taille fixe (3×3 , 7×7 ...)



Fenêtre de corrélation



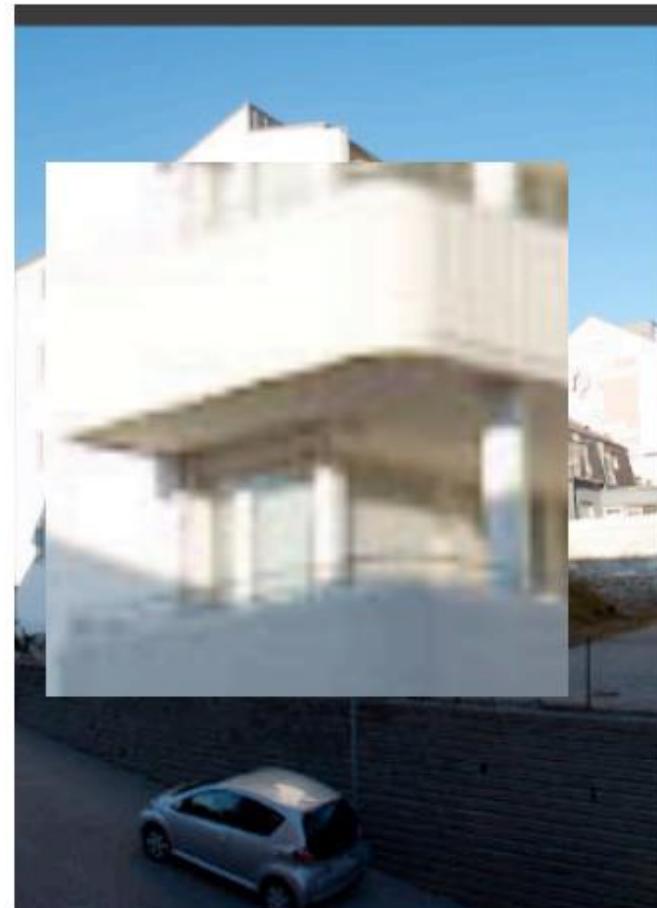
Example: **W** is a 3x3 window in red

w is a 9x1 vector

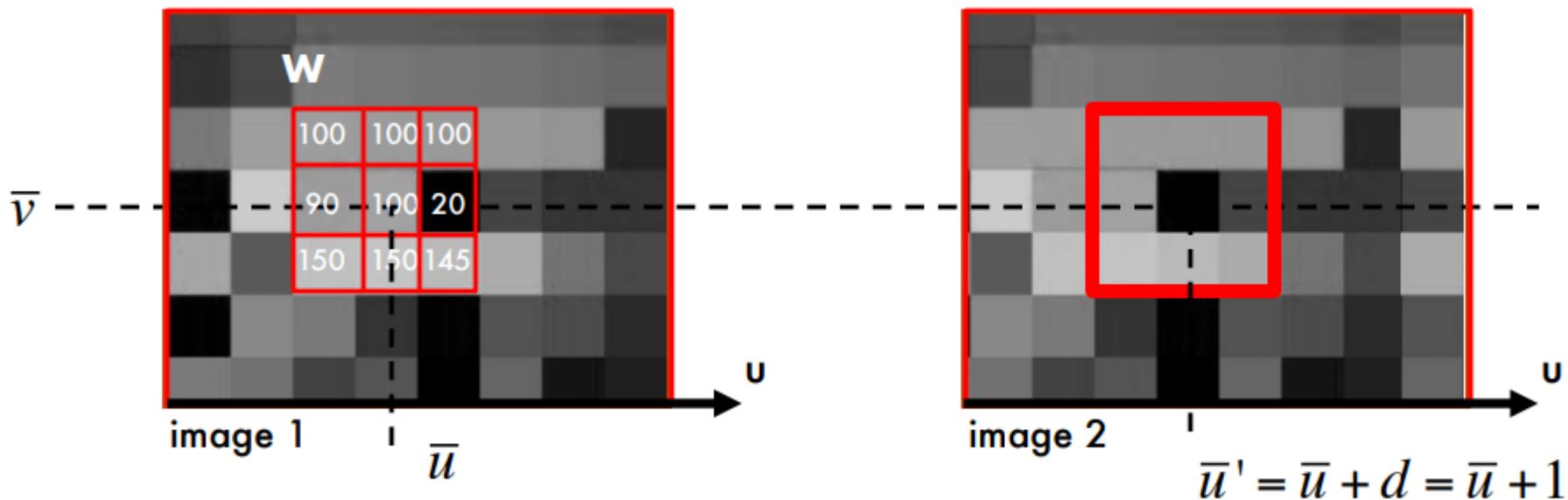
$$\mathbf{w} = [100, 100, 100, 90, 100, 20, 150, 150, 145]^T$$

- Pick up a window **W** around $\bar{p} = (\bar{u}, \bar{v})$
- Build vector **w**
- Slide the window **W** along $v = \bar{V}$ in image 2 and compute **w'** (u) for each u
- Compute the dot product $\mathbf{w}^T \mathbf{w}'(u)$ for each u and retain the max value

- Sensible aux différences d'exposition



Normalized cross-correlation



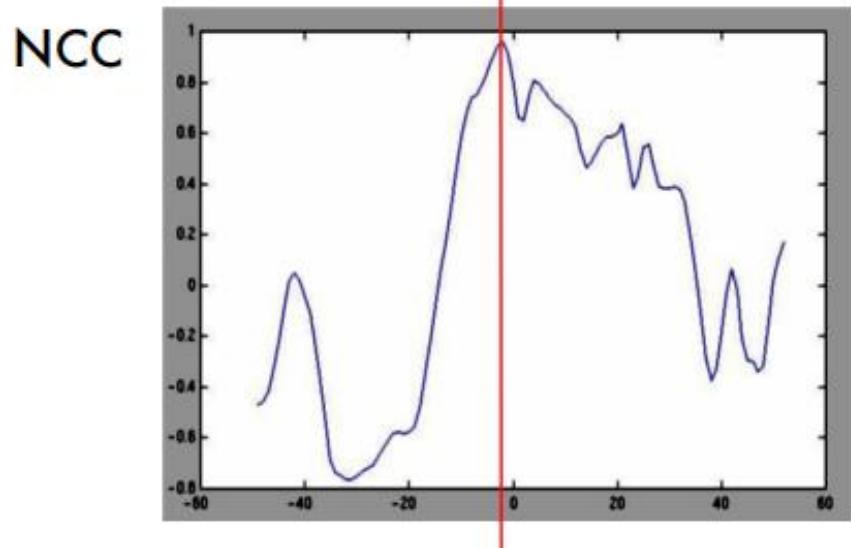
Find u that maximizes:

$$\frac{(w - \bar{w})^T (w'(u) - \bar{w}')}{\|(w - \bar{w})\| \| (w'(u) - \bar{w}') \|}$$

\bar{w} = mean value within w
located at u^{bar} in image 1

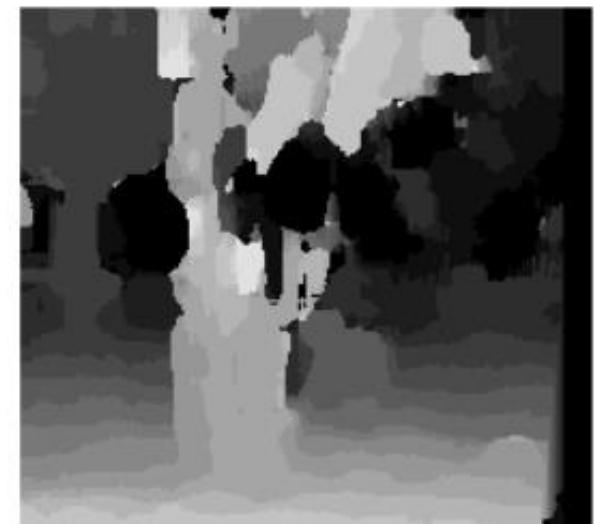
$\bar{w}'(u)$ = mean value within w'
located at u in image 2

Normalized cross-correlation



Taille de la fenêtre de corrélation

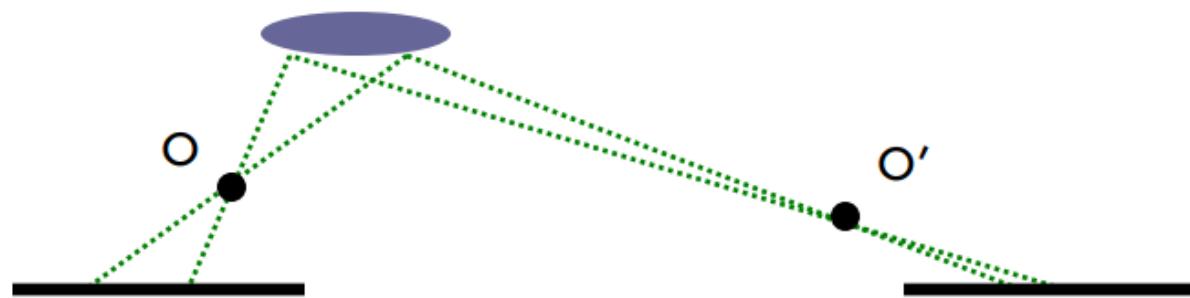
- Petite fenêtre:
 - + Plus de détail
 - Plus de bruit
- Grande fenêtre:
 - + Moins de bruit
 - Moins de détails
 - Problèmes sur les bords



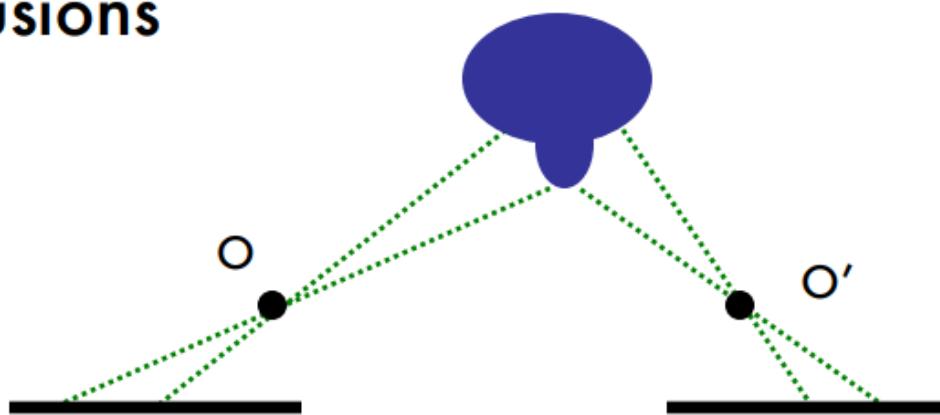
Window size = 3

Window size = 20

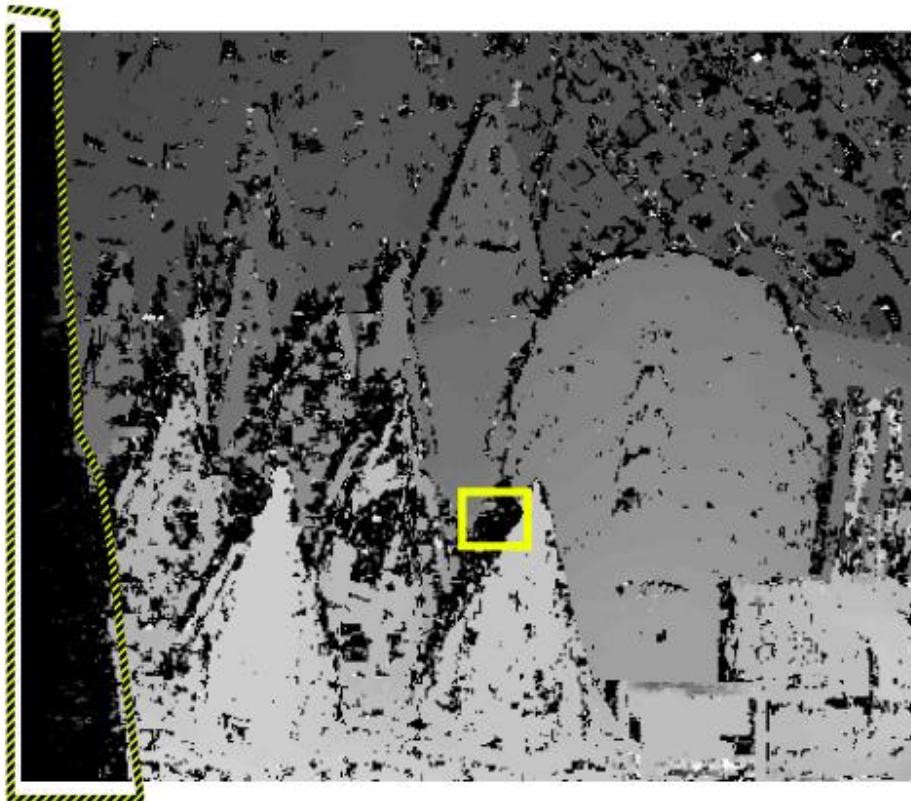
- Fore shortening effect



- Occlusions



Occlusions



Left image



Right image



- Régions homogènes et/ou peu texturées

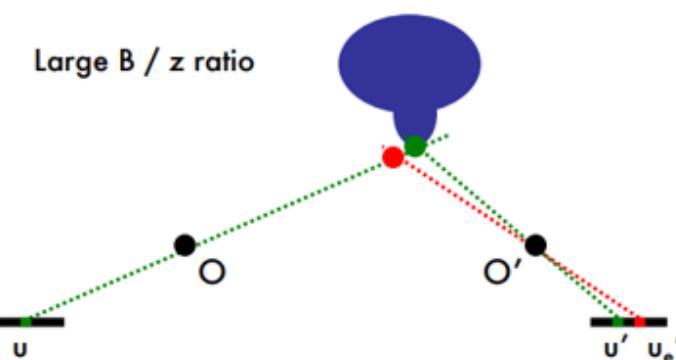
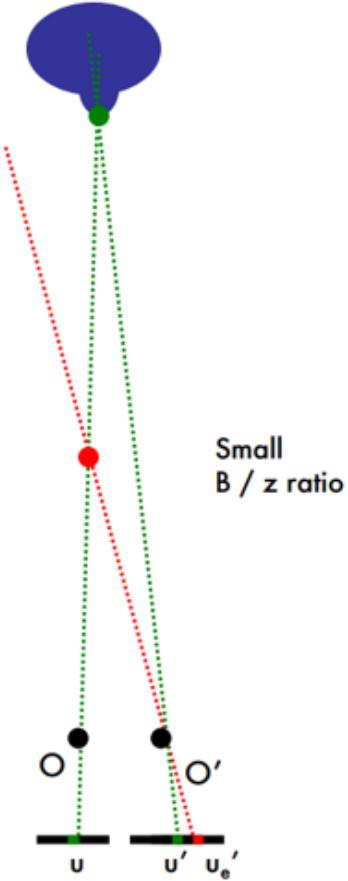


mismatch

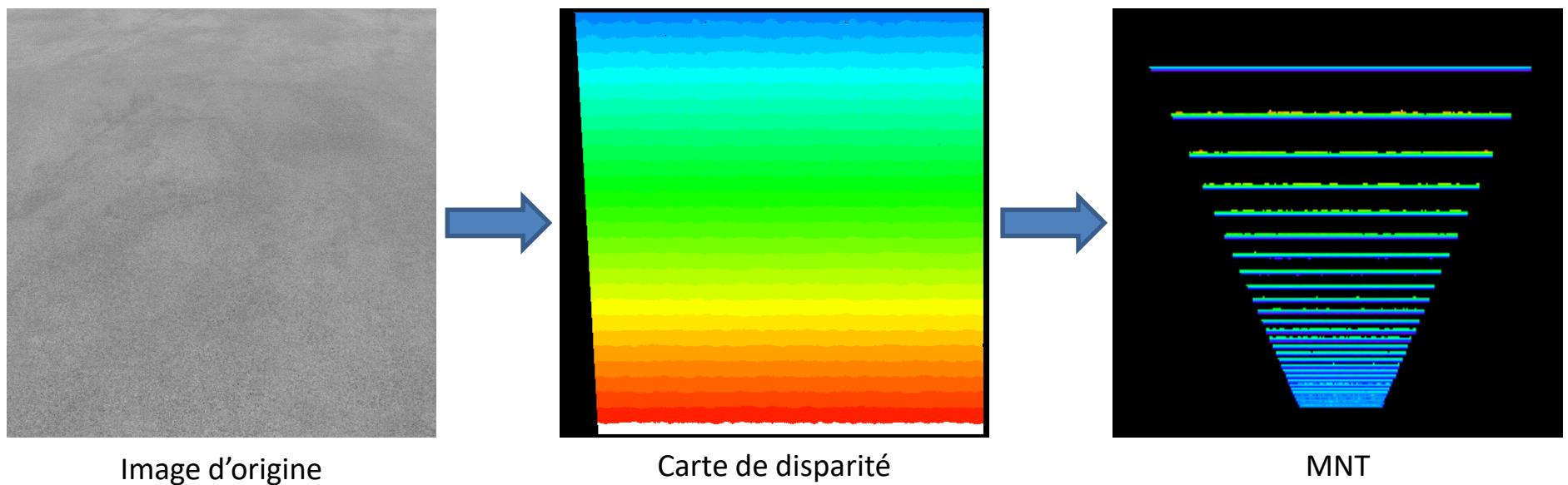
- Patterns répétitifs



- Petite baseline, i.e. petit $\frac{B}{z}$:
 - + Permet de voir des objets plus proches
 - + Moins d'occlusions
 - + Réduction des effets de foreshortening
 - Une petite erreur de mesure induit une grande erreur dans l'estimation de la profondeur => plus de bruit
- Grande baseline, i.e. grand $\frac{B}{z}$:
 - + Permet de voir des objets plus lointains
 - + Erreurs de mesure ont peu d'effet => moins de bruit
 - Plus d'occlusions
 - Impossible de voir les objets trop proches



- Le processus de stéréo-corrélation permet uniquement de calculer des valeurs entières de disparité.
 - Il est nécessaire de réaliser une interpolation pour lisser les disparités.
- Exemples sans interpolation sous-pixellique:

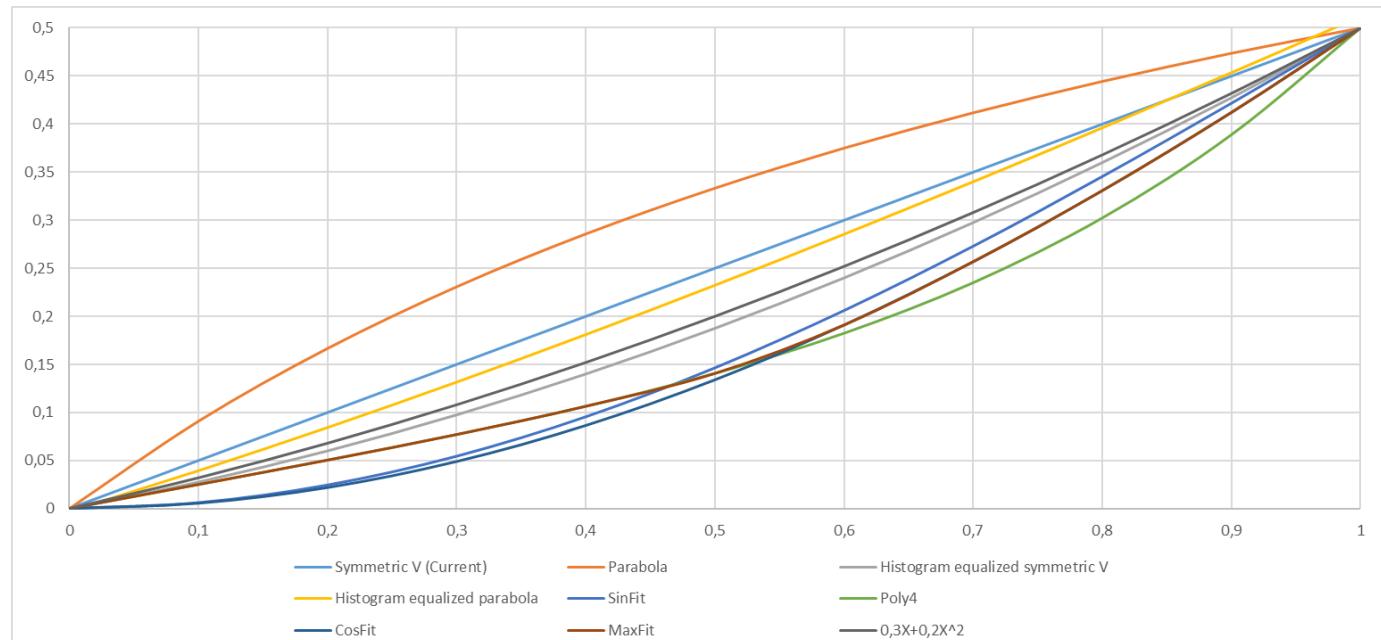


- Pour améliorer la précision d'un algorithme de stéréo-corrélation jusqu'à un niveau sous-pixellique, une fonction d'interpolation doit être implémentée, sa formulation déterminera la distribution des valeurs de disparité.
- La valeur de disparité est estimée en appliquant une fonction de la forme:
 - $d_{Subpix} = d_{int} + f(s_{left}, s_{med}, s_{right})$
 - Avec $s_{left}, s_{med}, s_{right}$ les scores de corrélation à gauche, au milieu et à droite.

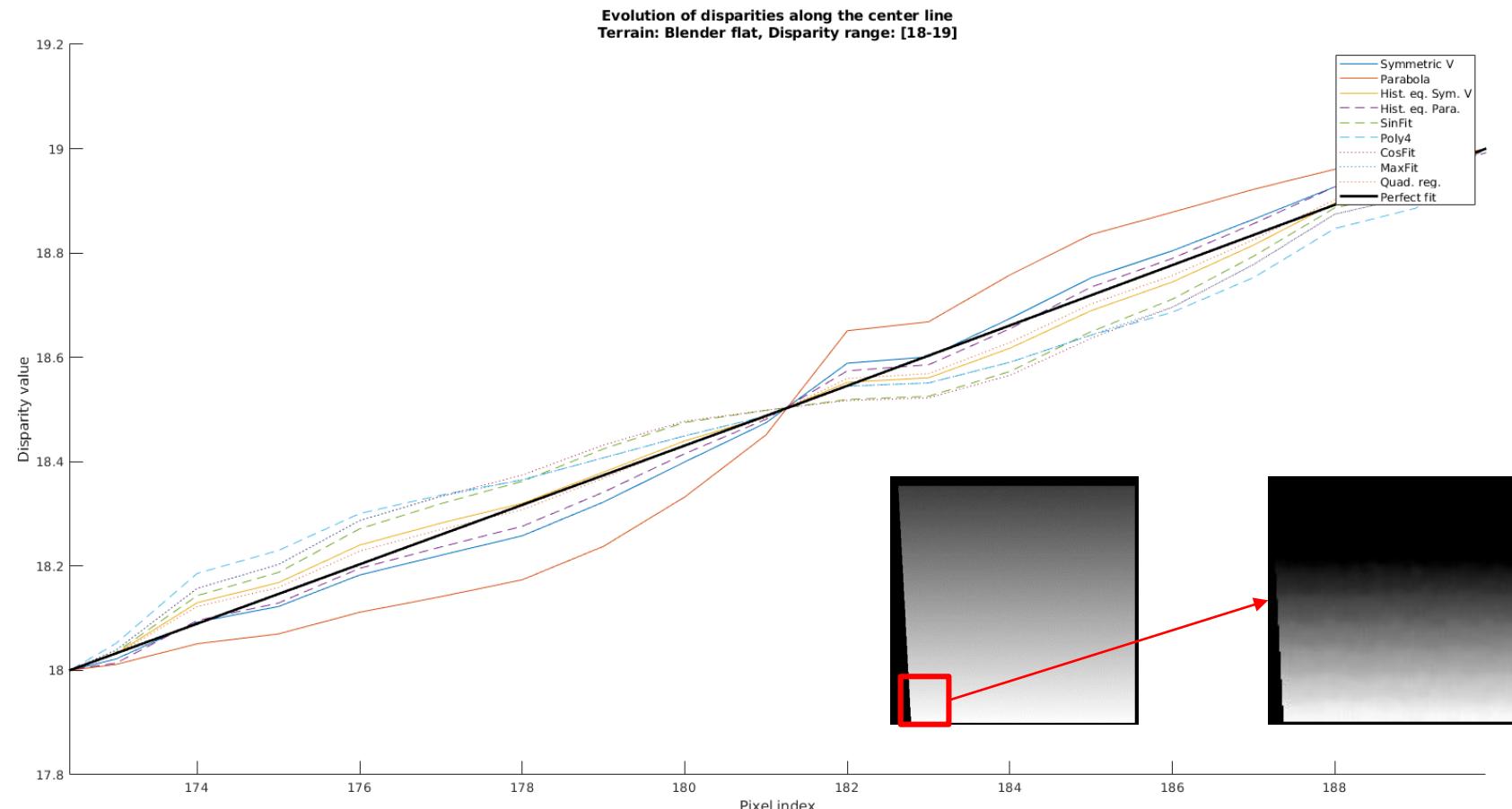
- La formulation de la fonction d'interpolation peut être simplifiée en:

$$d_{Subpix} = \begin{cases} d_{int} - 0.5 + f\left(\frac{l_d}{r_d}\right), & \text{if } l_d \leq r_d \\ d_{int} + 0.5 - f\left(\frac{r_d}{l_d}\right), & \text{otherwise} \end{cases}$$

- La fonction doit être monotone croissante et telle que:
 - $f: [0, 1] \rightarrow [0, 0.5]$

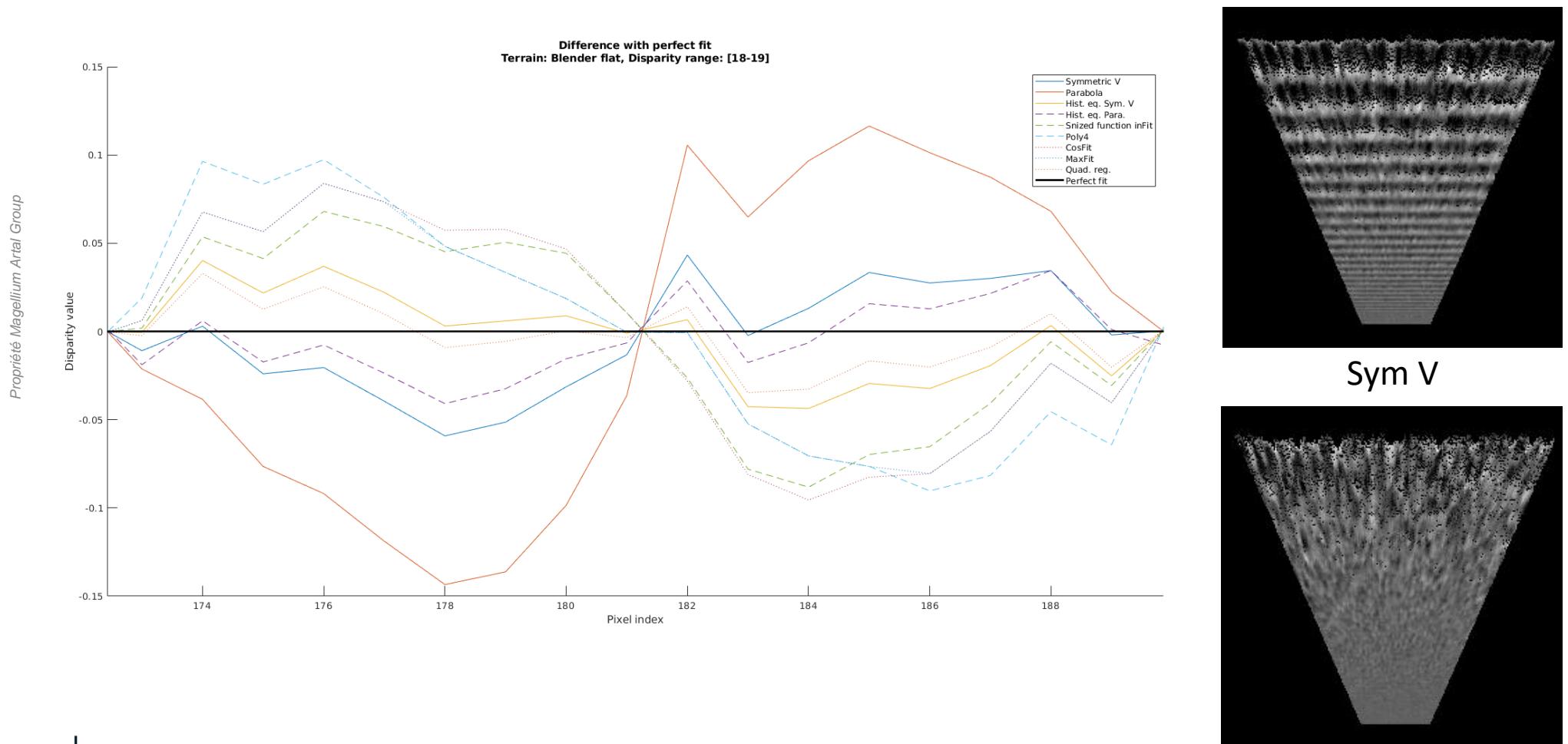


- L'utilisation d'une fonction d'interpolation sous-pixellique a pour effet d'introduire de petites erreurs sous la forme d'une sinusoïde:

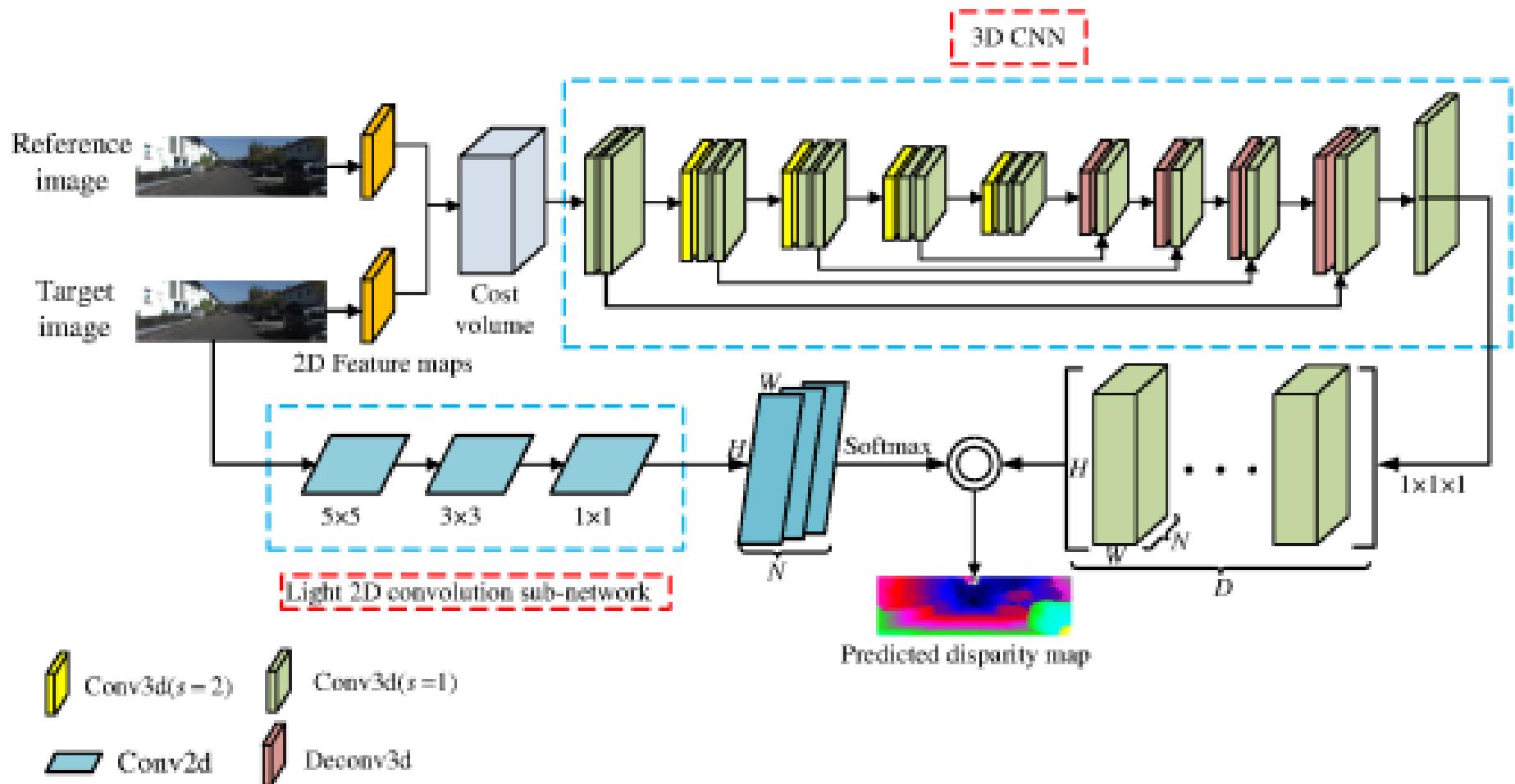


Effet de l'interpolation

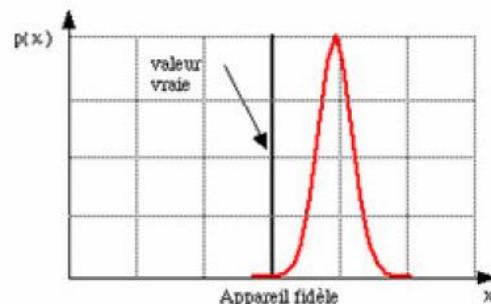
- Même si cet effet est difficilement visible sur la carte de disparité, il apparaît clairement sur la reprojeciton 3D:



- Réseaux de neurones convolutifs

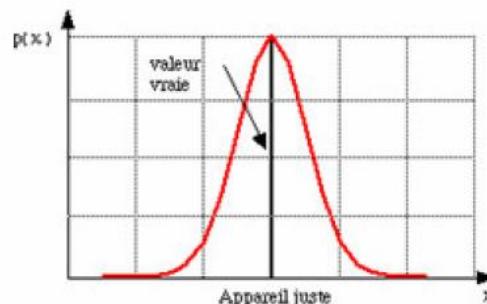


- Etat de l'art
 - <http://vision.middlebury.edu/stereo/>
- Evaluation



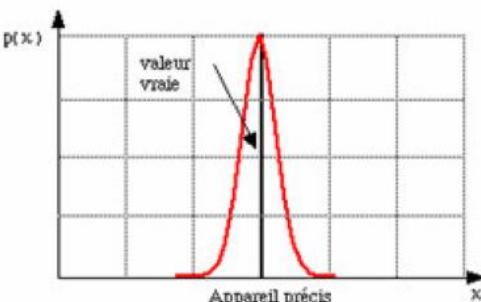
Fidélité : analysée par l'écart-type

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$



Justesse : analysée par la moyenne

$$\bar{x} = E(X) = \frac{1}{n} \sum_{i=1}^n x_i$$



Précision : analysée par la valeur RMS

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - x_{vrai})^2}$$

Merci pour votre attention