

VPOA

Raphaël Viards

Airbus Operations SAS

raphael.viards@airbus.com

Introduction



Plan du cours

I

Perception 3D

1. Introduction
2. Systèmes d'acquisition 3D
3. Modélisation de la camera
4. Stéréovision

II

Analyse de l'environnement

1. Introduction
2. Analyse de surfaces et Reconstruction 3D
3. Segmentation 3D
4. Recalage 3D

III

Détection et Localisation

1. Introduction
 2. Géométrie projective
 3. Extraction de caractéristiques locales
 4. Reconnaissance d'objets
 5. Estimation du mouvement
- 

Qu'est-ce que la computer vision?



[www.menti.com](https://www.menti.com/poll/89678448) code 8967 8448

Qu'est-ce que la vision par ordinateur?

40 réponses

extraction de features
acquisition d'images
traitement d'image classification
ia detection les yeux de l'ordinateur
machine learning
détection
segmentation reconnaissance
deep learning
intelligence artificielle
traitement
automatisation
gpu
analyse
pipeline donnée
capteurs
captation modélisation
distinguer ce que voit l'ordinateur
detection instances class
représentation réaliste

Computer Vision

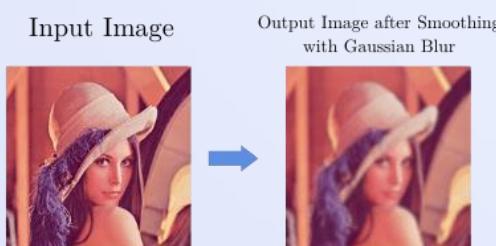
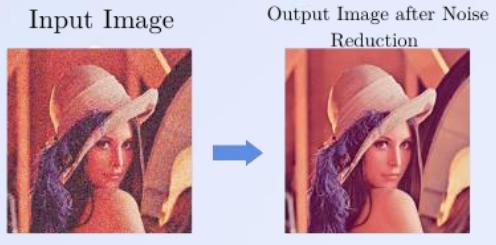
La **computer vision** est un domaine interdisciplinaire dont le principal but est de permettre à une machine d'**analyser**, **traiter** et **comprendre** une **représentation** de l'environnement obtenue par un système d'**acquisition**.

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



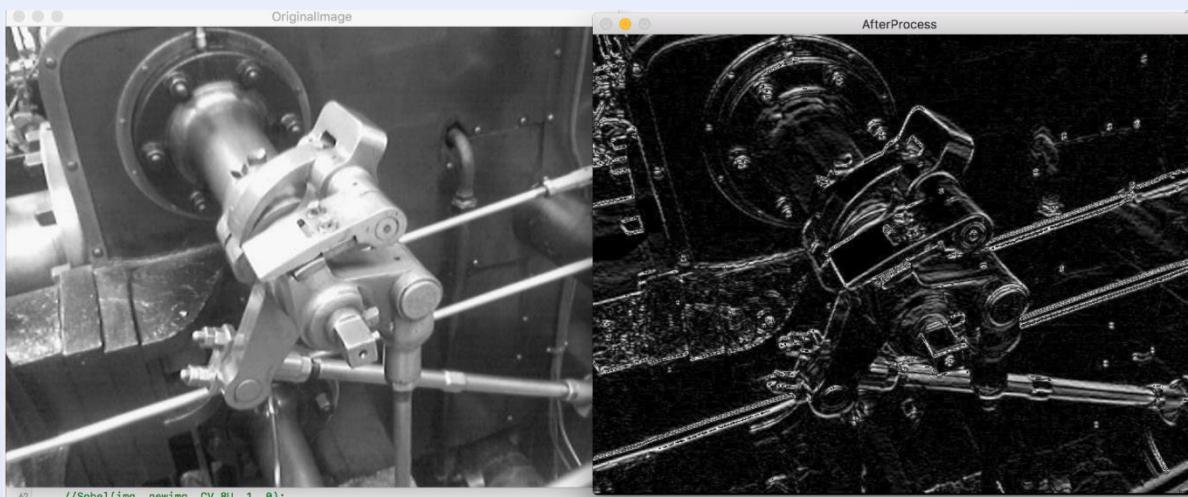
Image Processing

On peut assimiler le **traitement d'image** à une boîte prenant une **image en entrée**, la transformant en interne, et renvoyant une nouvelle **image en sortie**.



Le traitement d'image peut inclure de **nombreuses techniques** différentes. Le résultat final peut ensuite être utilisé pour une analyse plus approfondie ou une prise de décision.

Il s'agit d'un processus crucial pour de nombreux domaines. Généralement, le traitement d'image correspond à la **première étape** de nos applications.



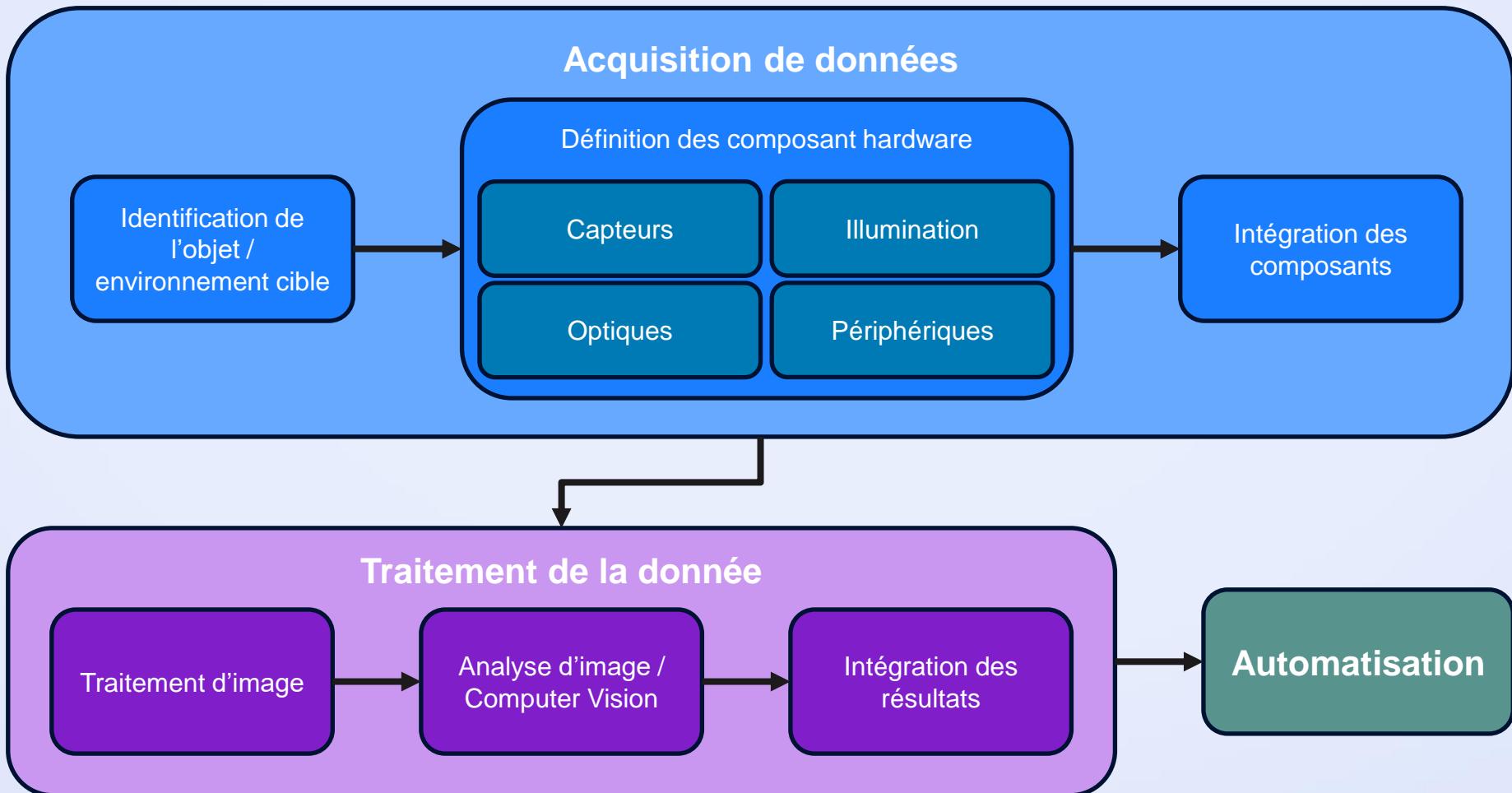
Machine Vision

La **machine vision** correspond aux **technologies** et **méthodes** utilisées pour fournir des **solutions automatisées** d'inspection et d'analyse d'images pour des applications telles que l'inspection automatisée, le contrôle de procédés ou encore le pilotage de robots, principalement dans l'**industrie**.

La machine vision fait référence à de nombreuses technologies et produits **software ou hardware**, à des systèmes intégrés, des actions, des méthodes et expertises. La machine vision est une discipline d'**ingénierie des systèmes** qui peut être considérée distincte de la computer vision, une forme de « computer science ».

	Machine Vision	Computer Vision
Discipline	Ingénierie des systèmes	Computer science, informatique
Objectif	Résoudre des problèmes spécifiques	Reproduire les capacités de perception humaines
Domaine d'application	Automatisation industrielle	Laboratoires, recherche académique
Processus	Compréhension du monde, sélection hardware, acquisition de données, traitement, automatisation	Démarre au niveau de la conception algorithmique

Vision par ordinateur



Perception 3D

Introduction



Perception

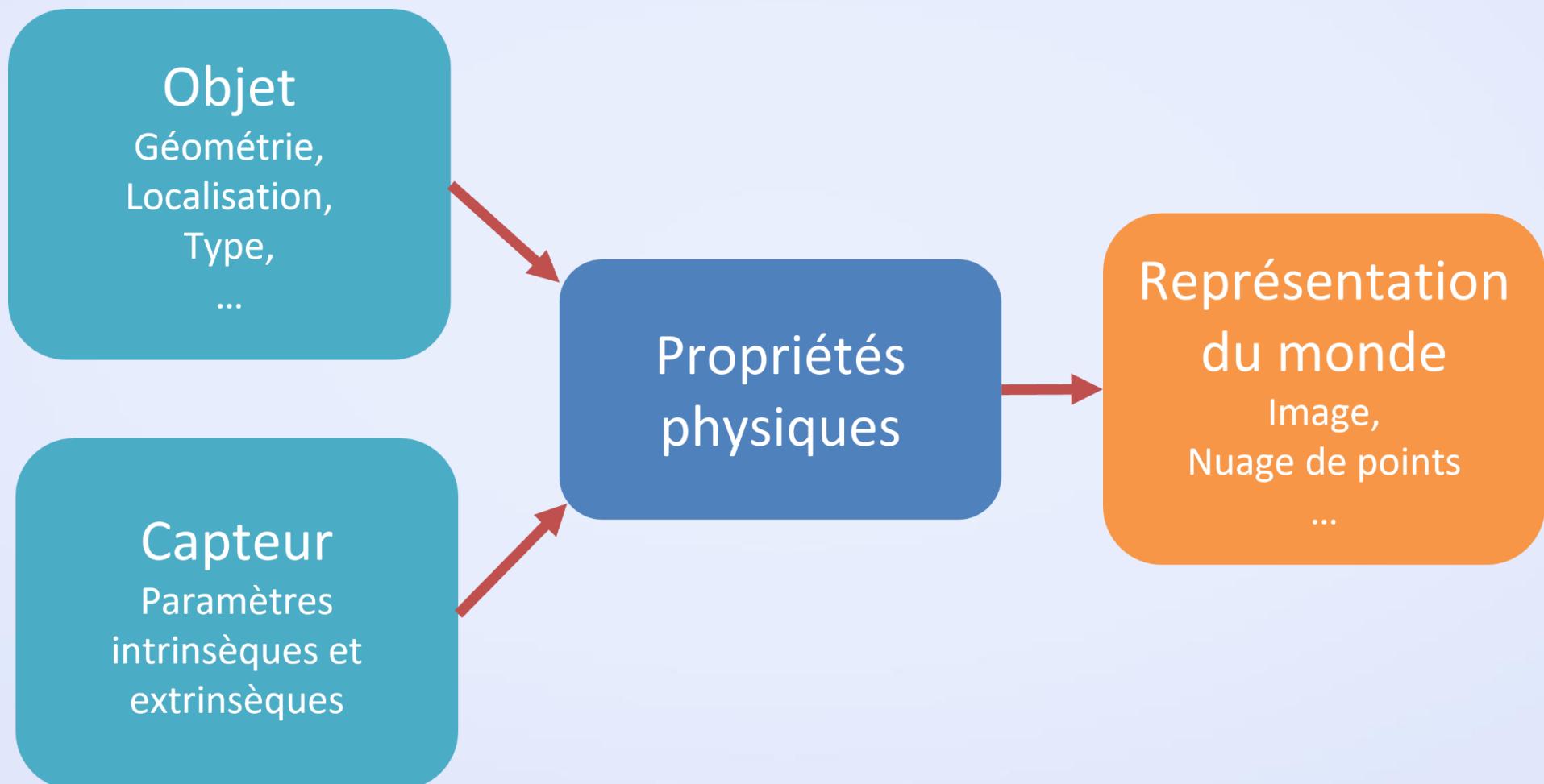
Qu'est-ce que la perception?

- **Estimation** des propriétés **géométriques** et **sémantiques** d'un ou plusieurs objets, à partir d'**observations** produites par un **système d'acquisition**.

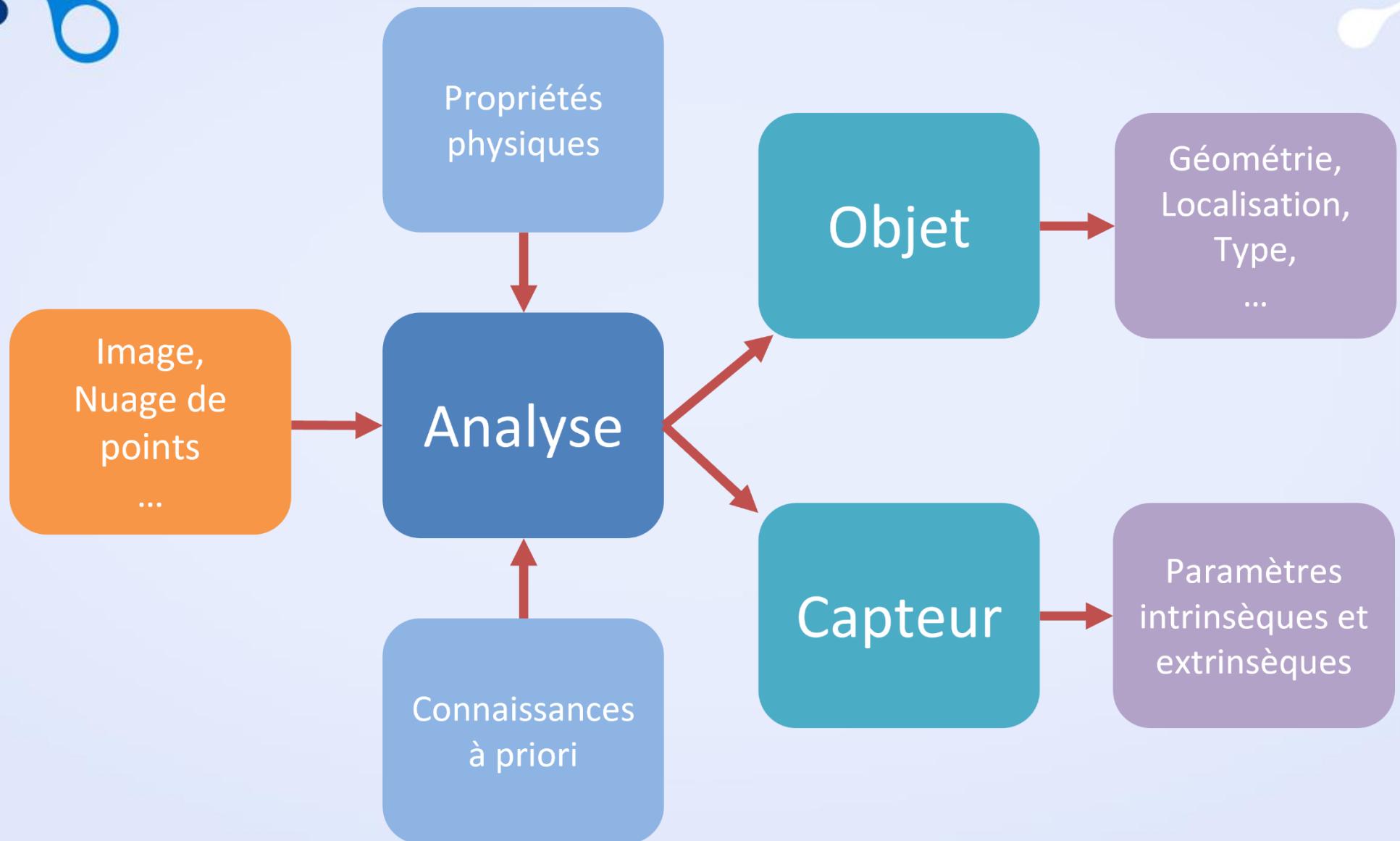
Que peut-on obtenir?

- 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
- ...

De l'objet à l'observation



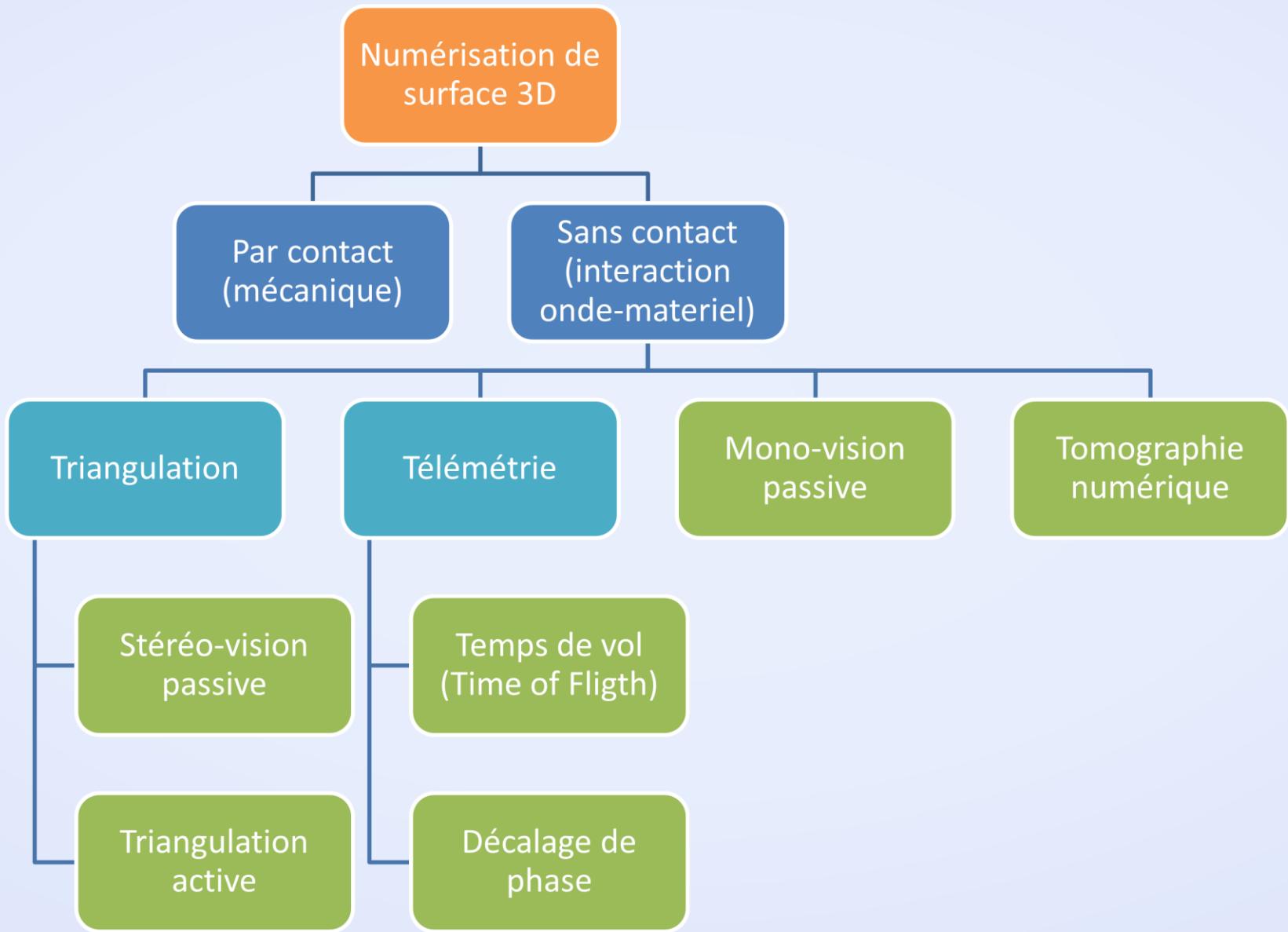
De l'observation à l'objet





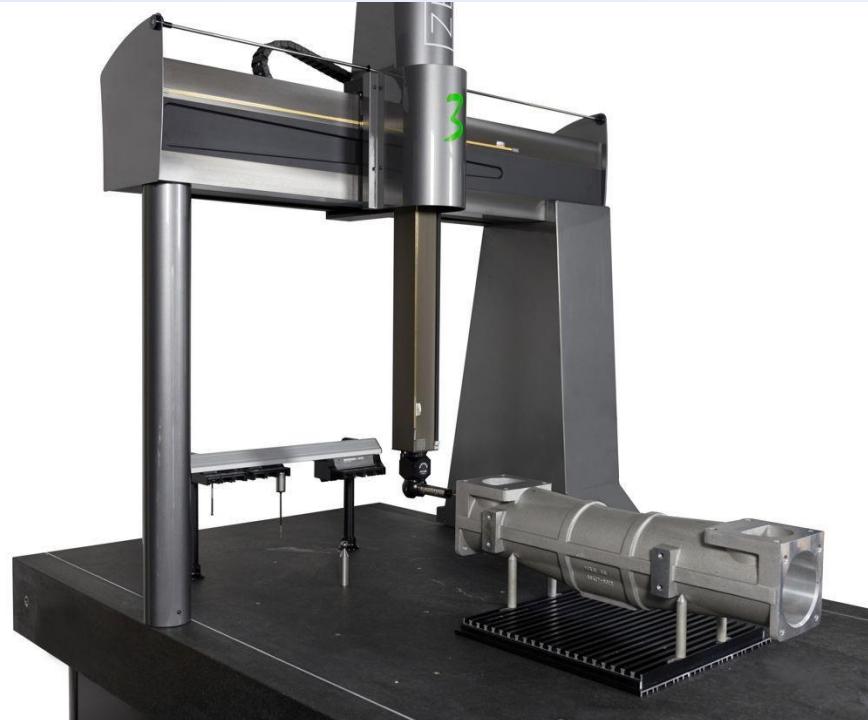
Systèmes d'acquisition 3D

Systèmes d'acquisition 3D





Numérisation par contact



Machine à mesurer tridimensionnelle

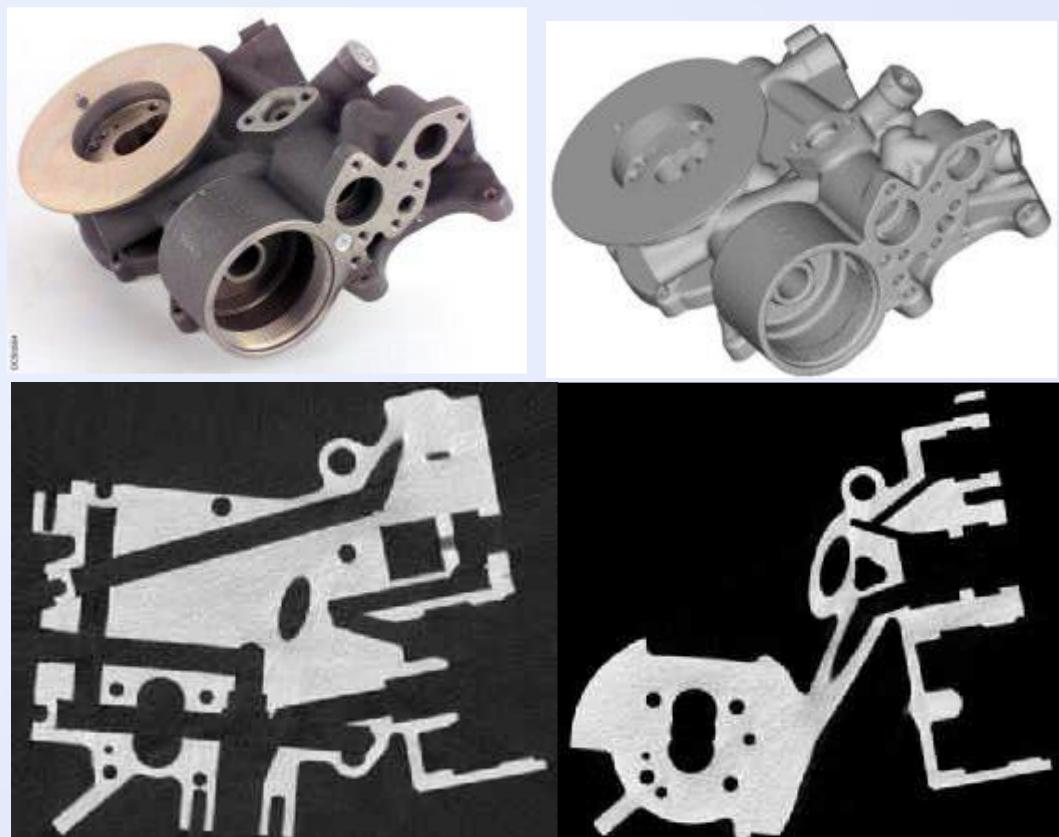
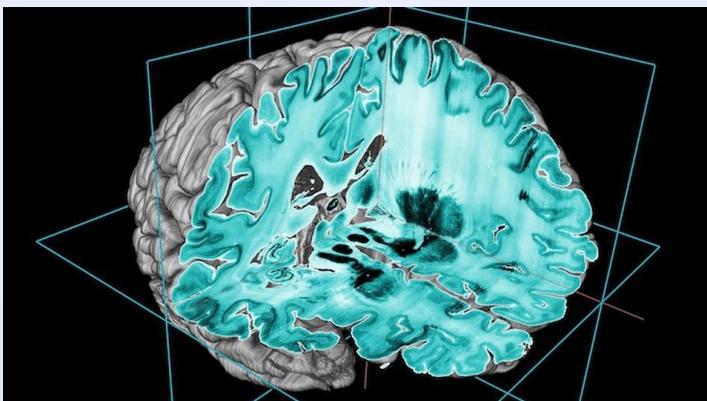


Bras articulé



Tomographie numérique

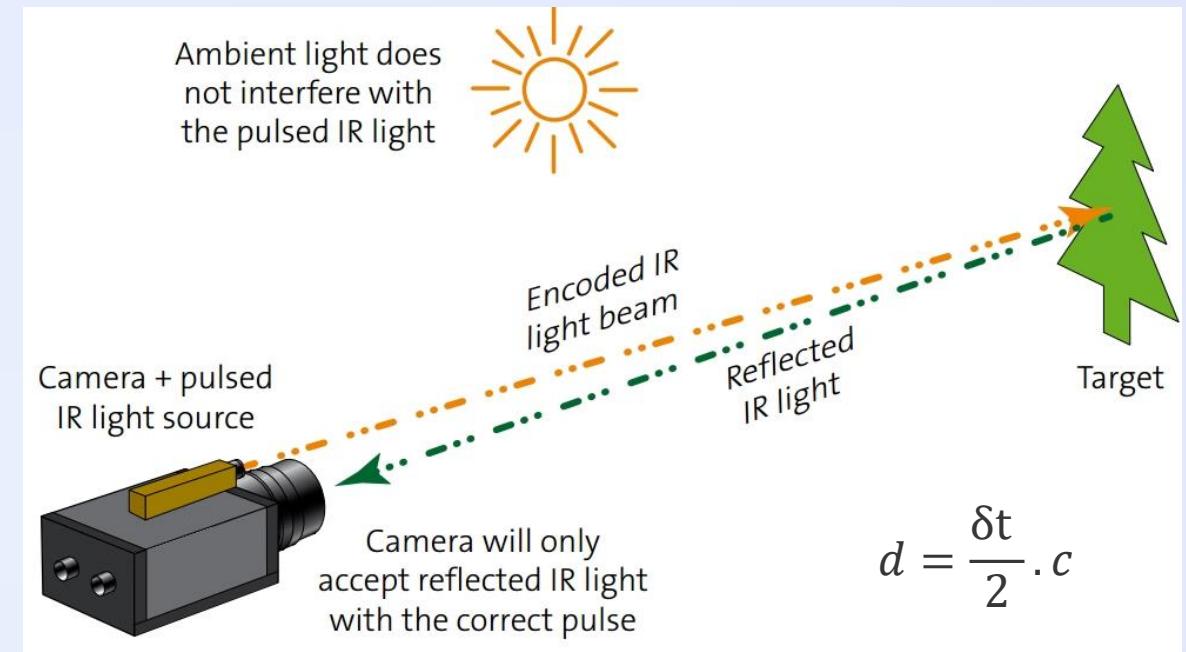
- Rayons X
 - Principe de l'IRM
 - Superposition de coupes 2D successives





Télémétrie par temps de vol

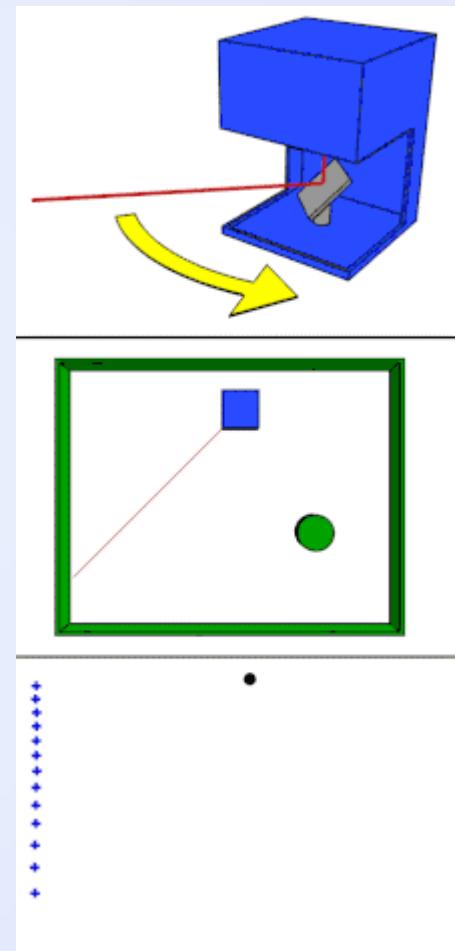
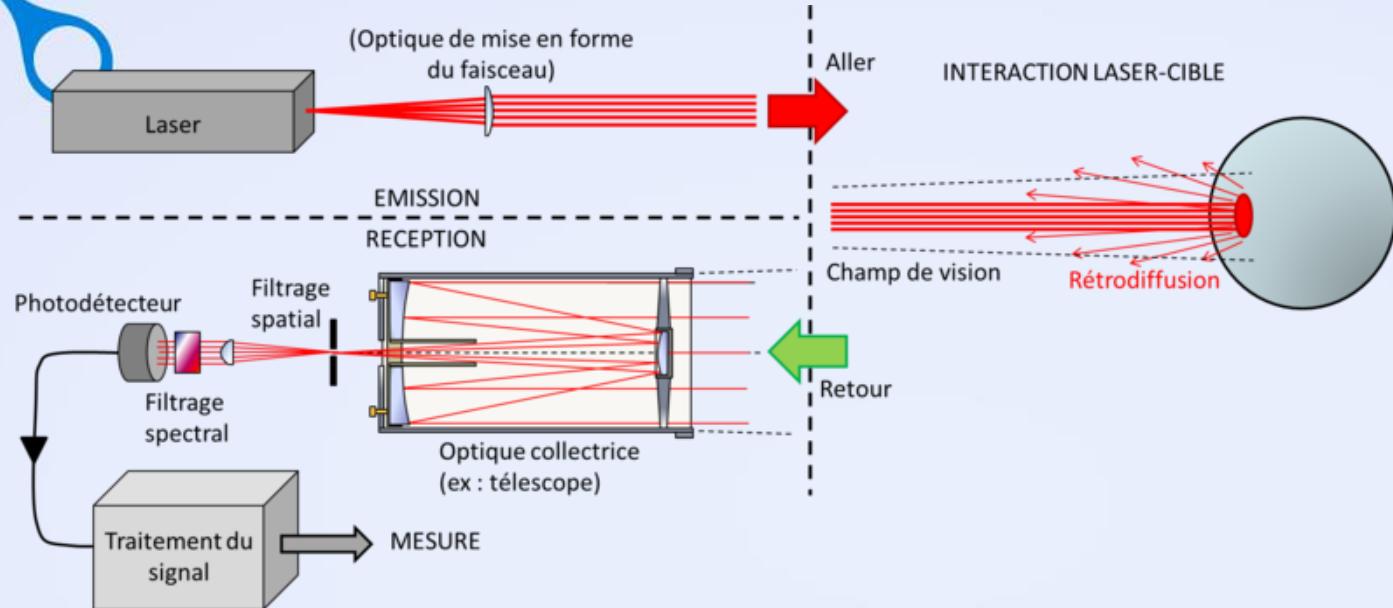
- Envoi d'une impulsion pulsée:
 - Lumineuse classique: Caméra ToF
 - Lumineuse Laser : LIDAR
 - Ultrasonore : SONAR
 - Ondes radio : RADAR



- Mesure du temps de vol (ToF: Time of Flight)

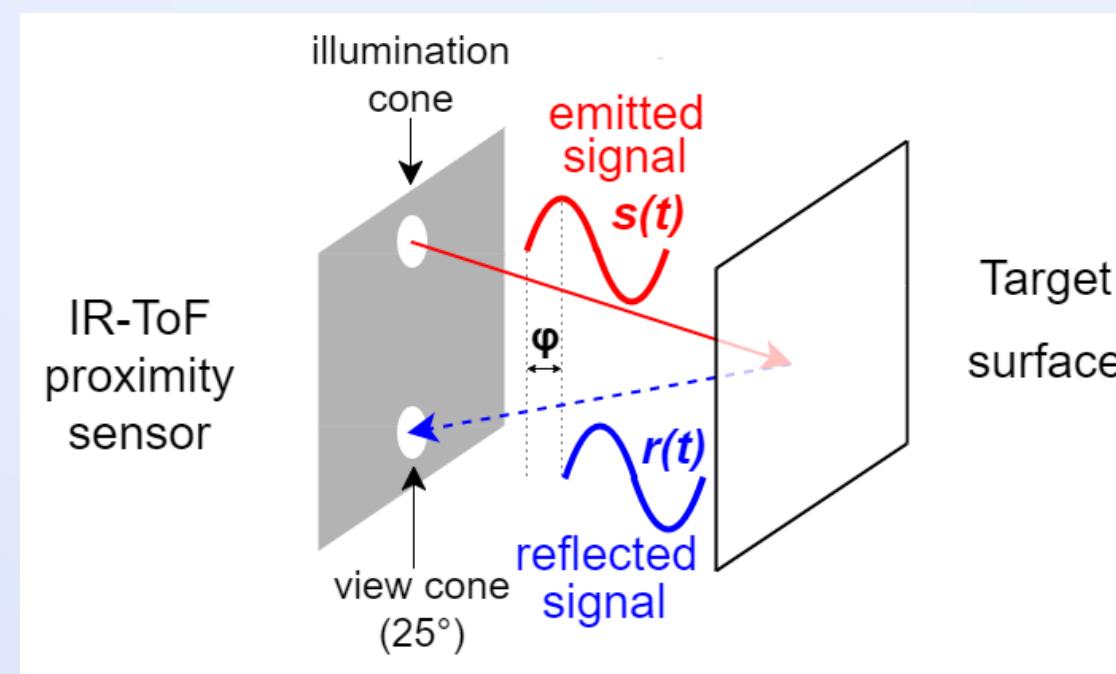


Lidar ToF



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

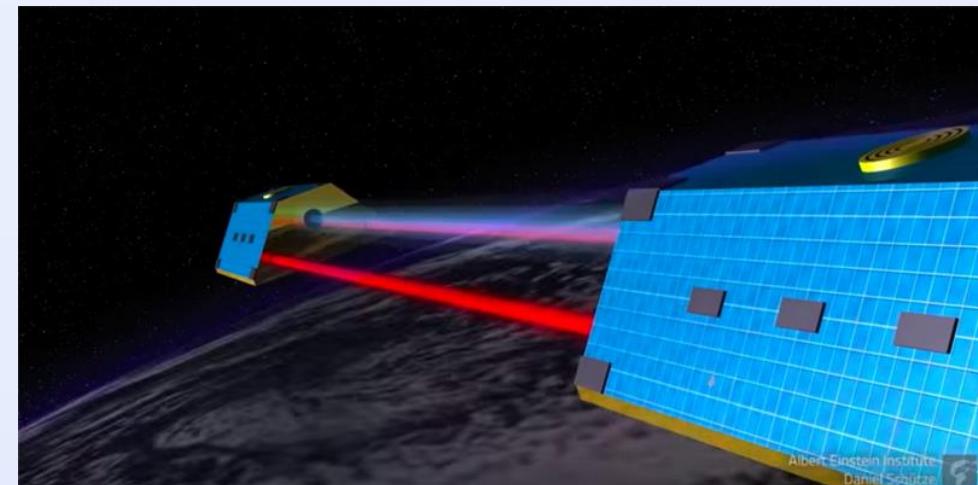
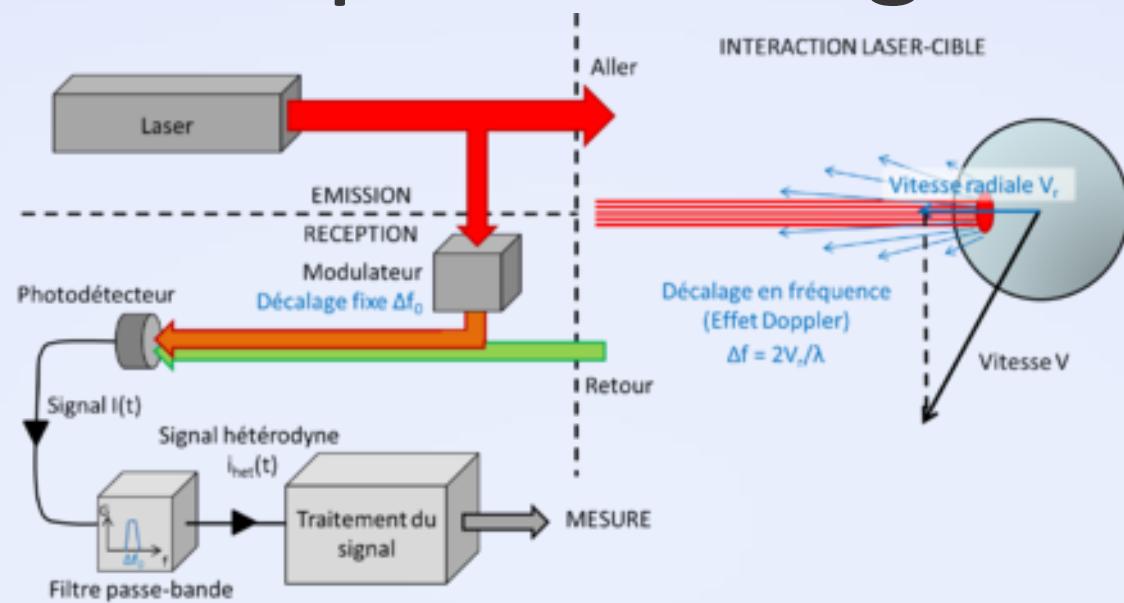


$$\varphi = \delta t \cdot 2\pi \cdot f$$

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

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

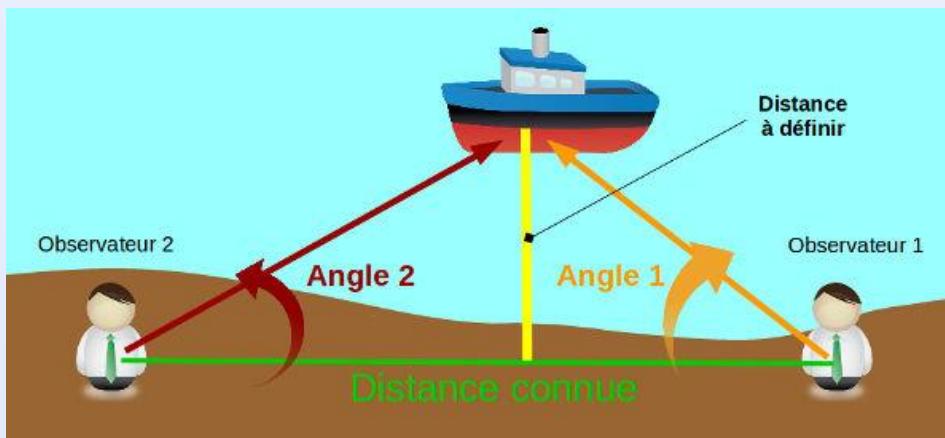
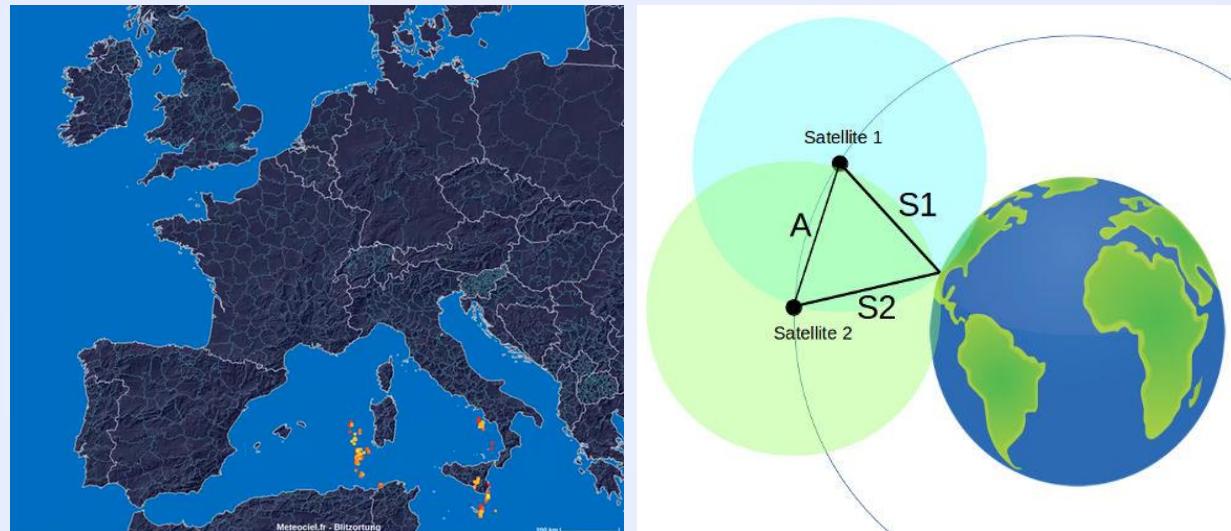
Télémétrie par décalage de phase





Triangulation

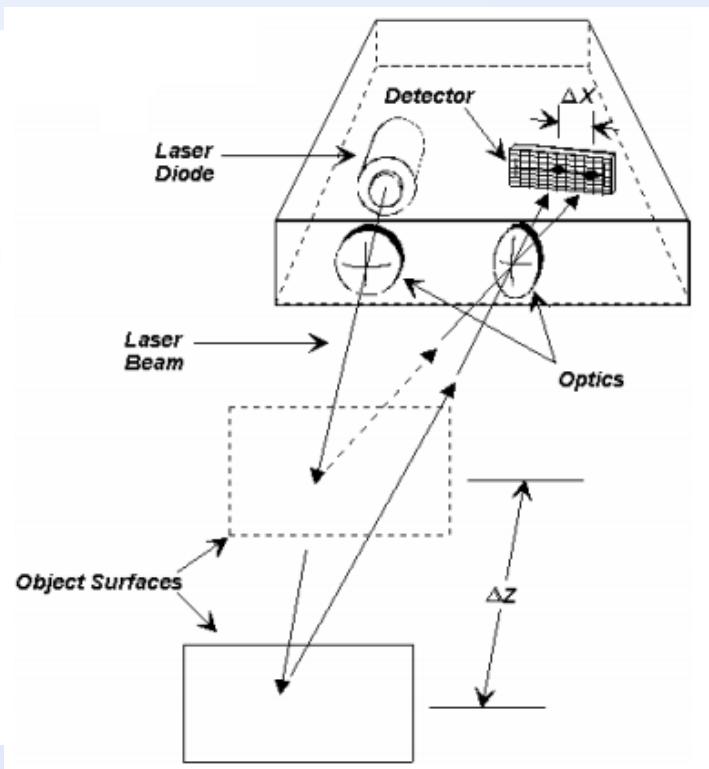
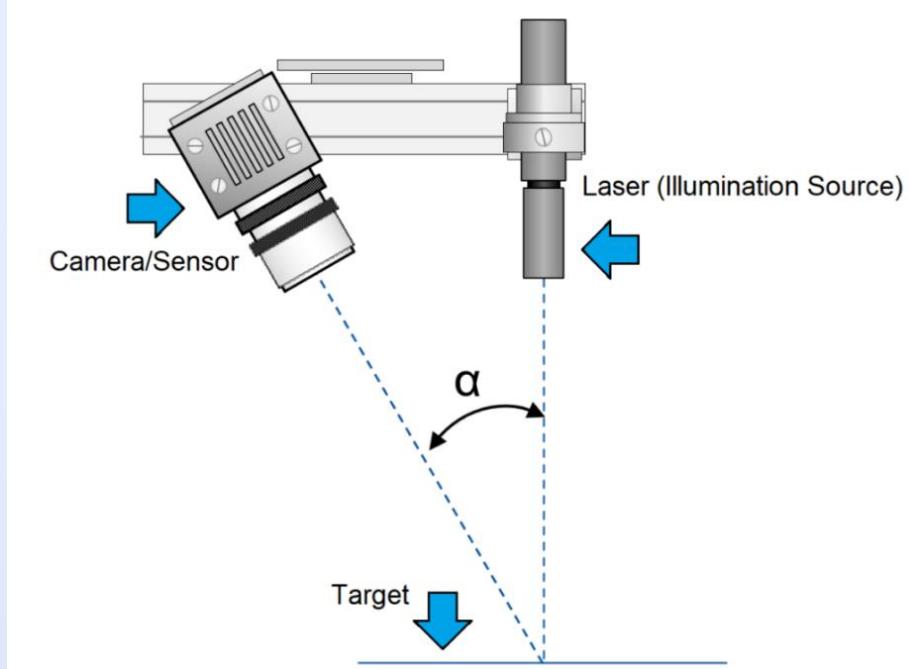
Principe utilisé par les géomètres, le GPS...



Triangulation active

Pointeur laser

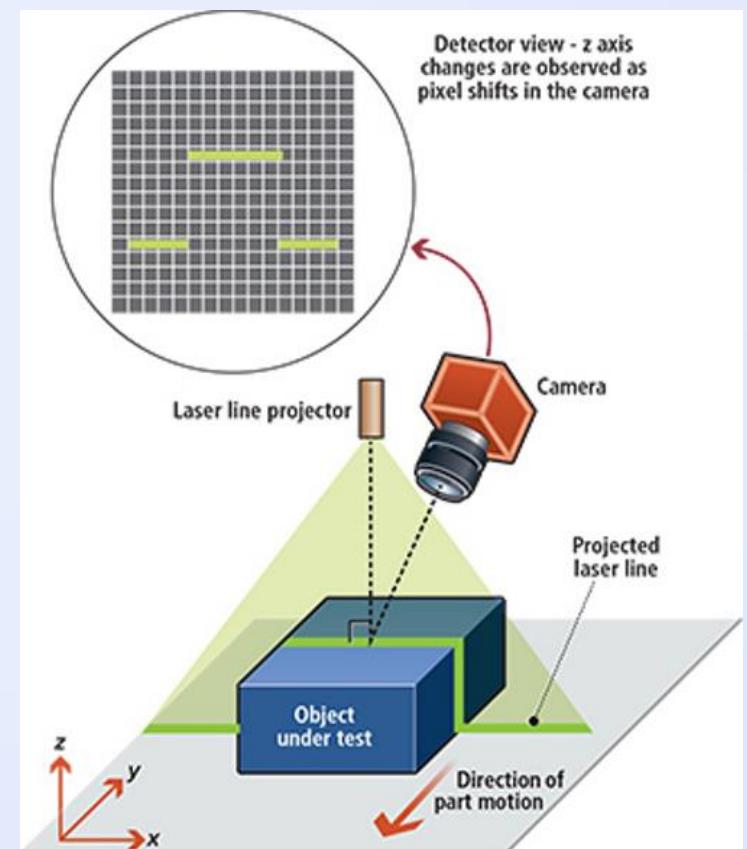
- Un rayon laser est envoyé vers l'objet à mesurer
- La lumière diffusée est observée par une camera
- On peut en déduire la profondeur du point



Triangulation active

Profilométrie:

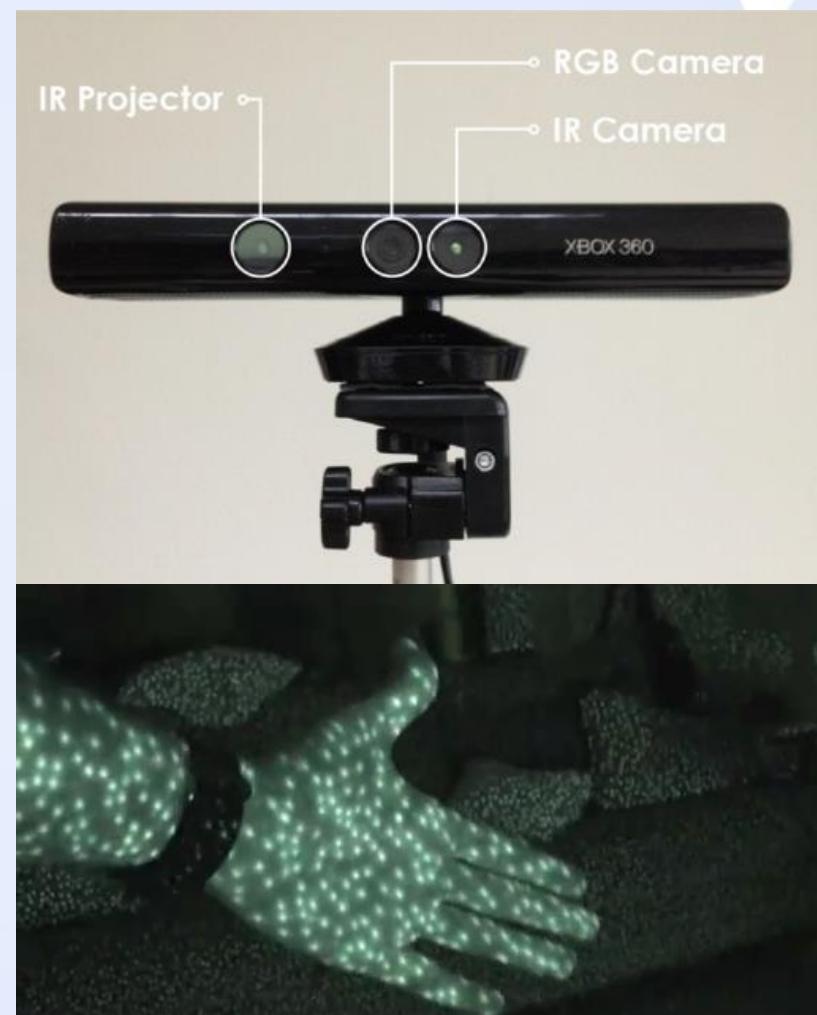
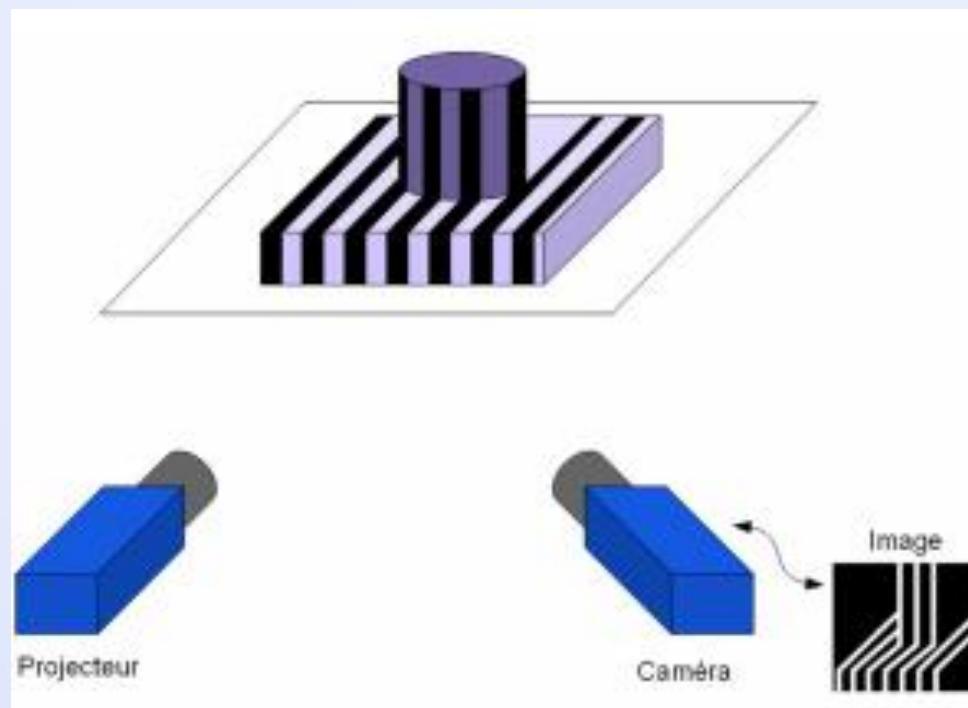
- Projection d'une ligne à l'aide d'un laser
- Analyse de la déformation de la ligne dans l'image
- Un unique balayage suffit pour assurer la couverture de la surface



Triangulation active

Lumière structurée:

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





Caméras

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



Caméras

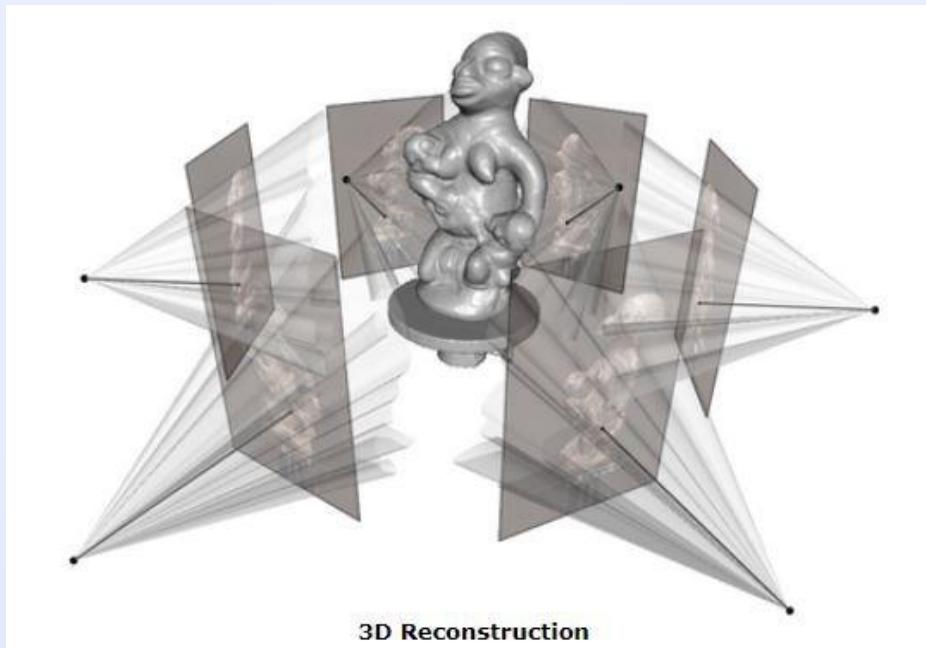
Inconvénients:

- 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



Mono-vision passive

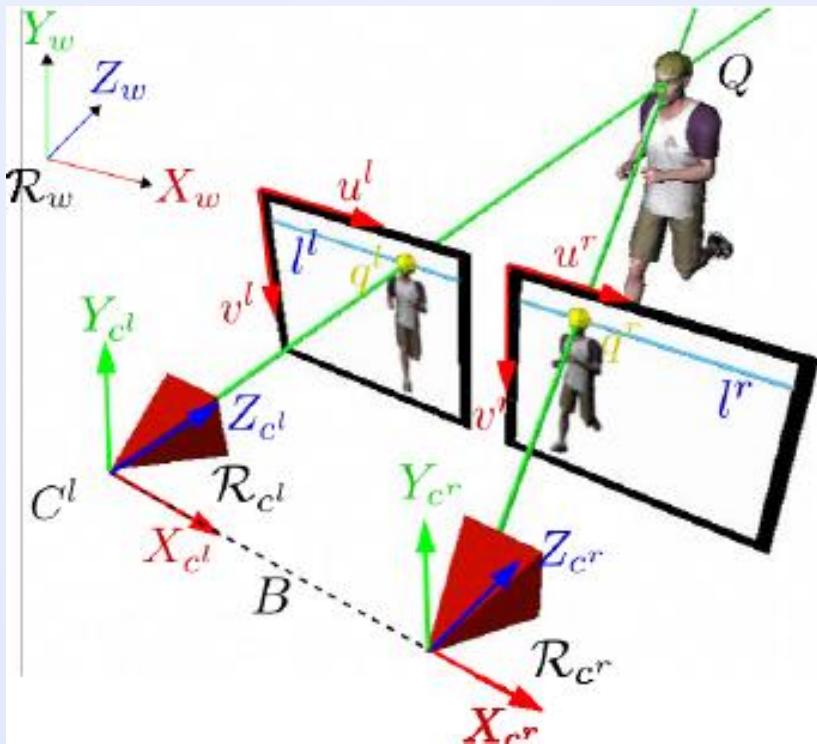
- Utilisation de simples images 2D pour obtenir de l'informations 3D
- Obtention d'angles de vue différents par mouvement de l'objet et/ou de la caméra



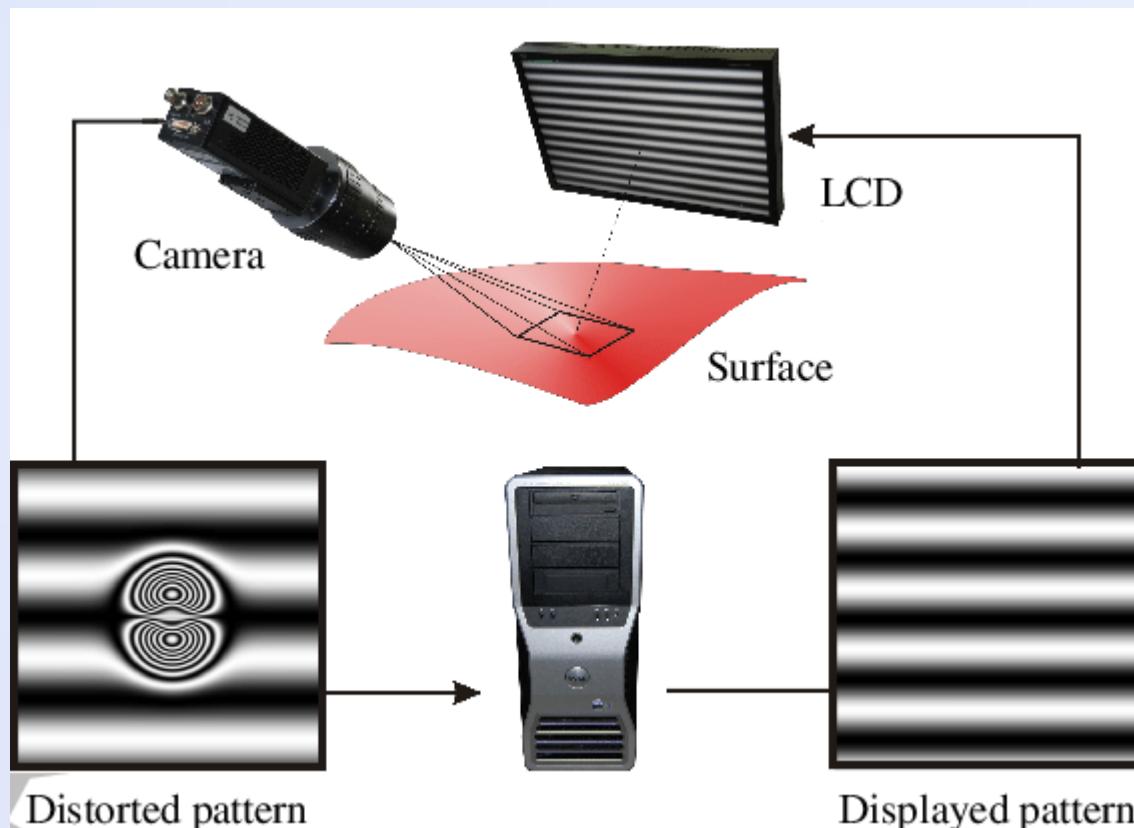


Stéréovision

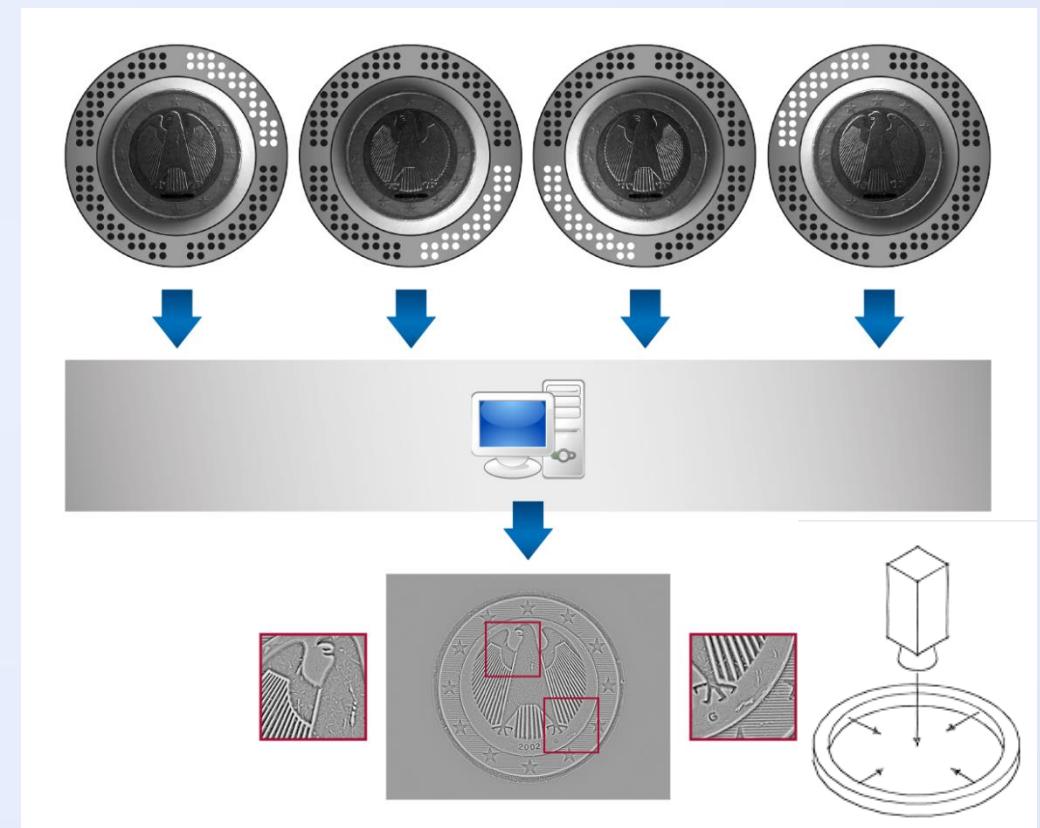
- Vision d'une même scène depuis deux endroits légèrement décalés l'un par rapport à l'autre
- Principe de la perception 3D chez l'Homme



Autres



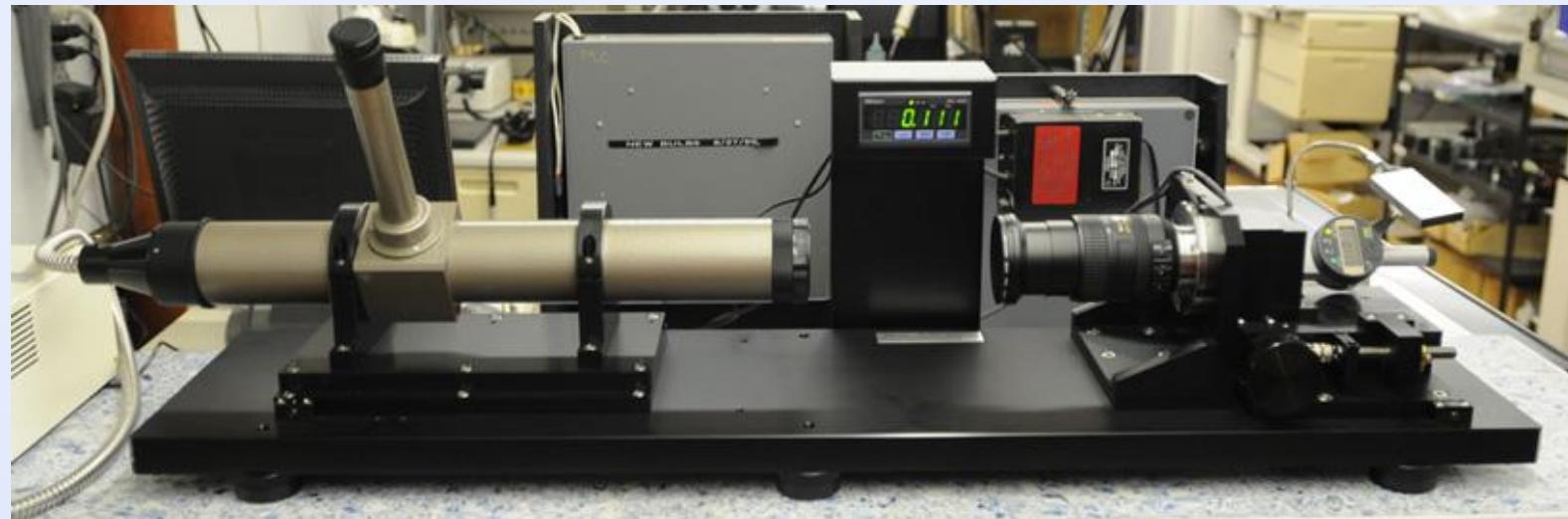
Deflectometrie



Shape from shading

Précision et étalonnage

- 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





3

Modélisation de la camera

Géométrie optique

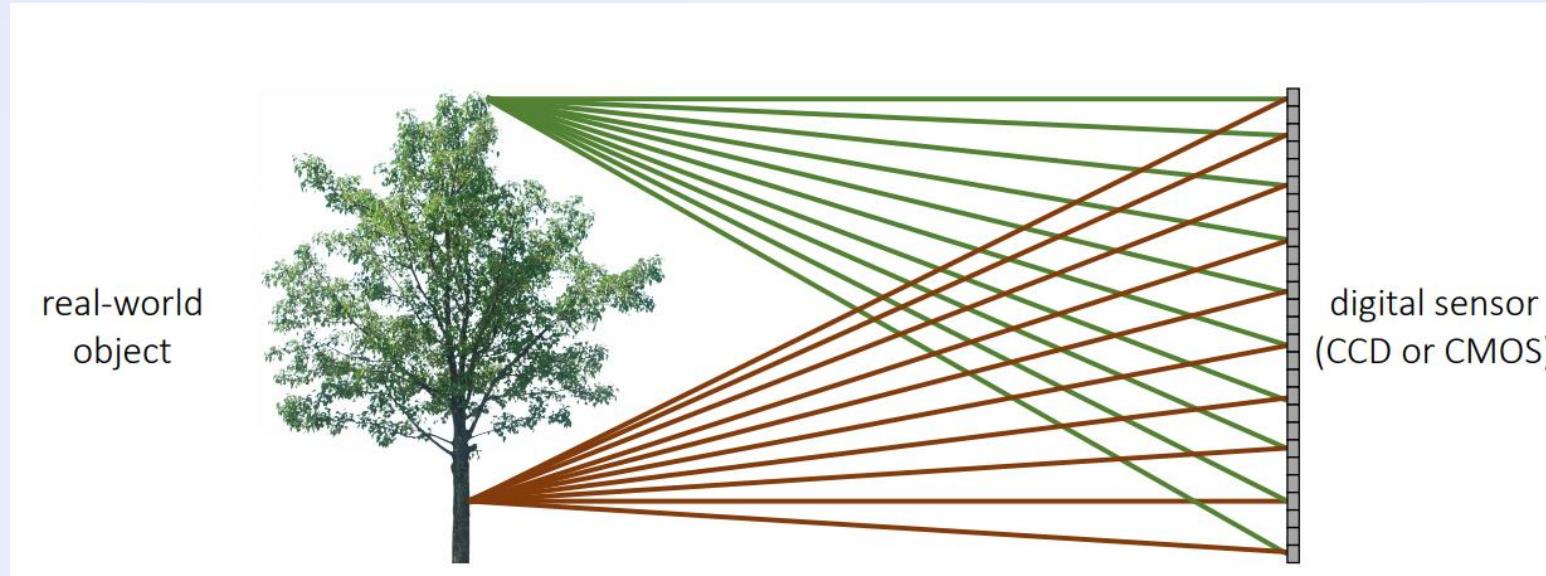
Postulats:

- 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



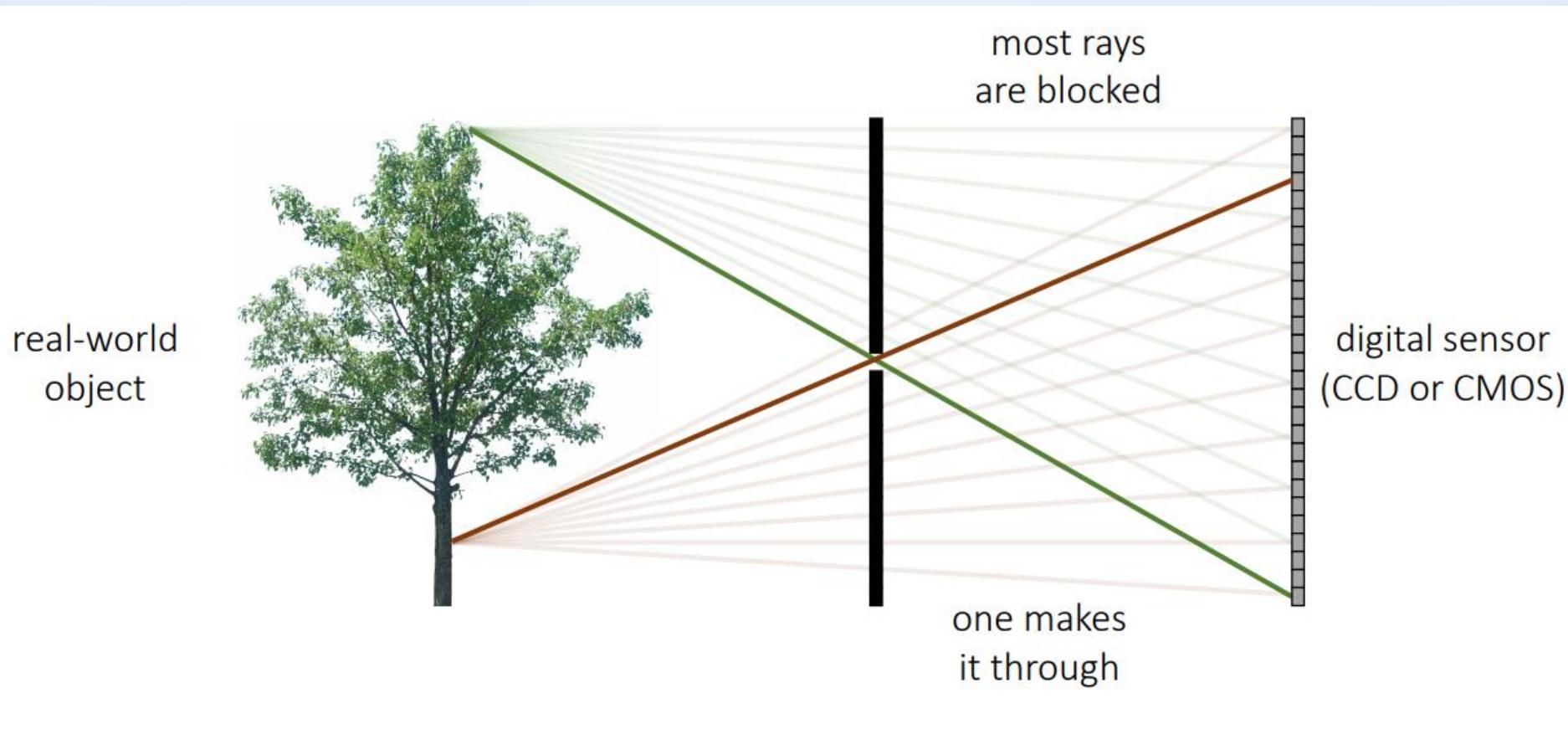
Modélisation de la camera

Capteur uniquement

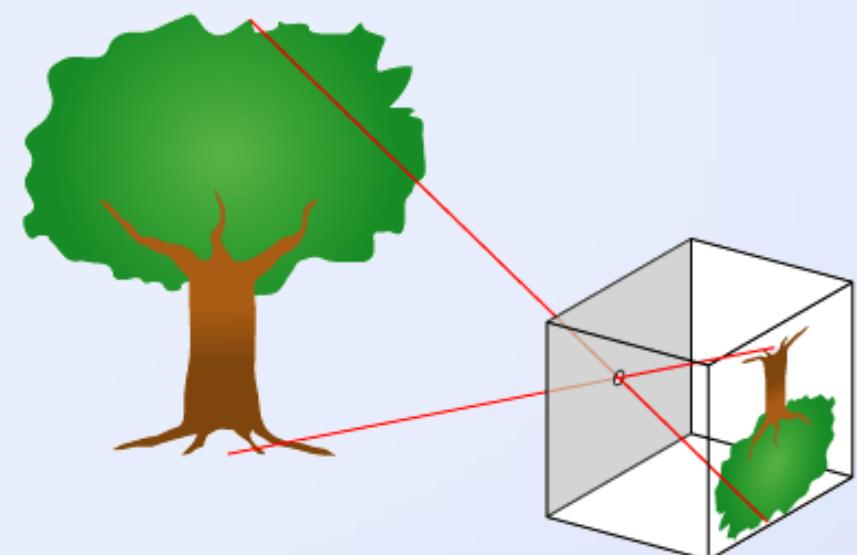
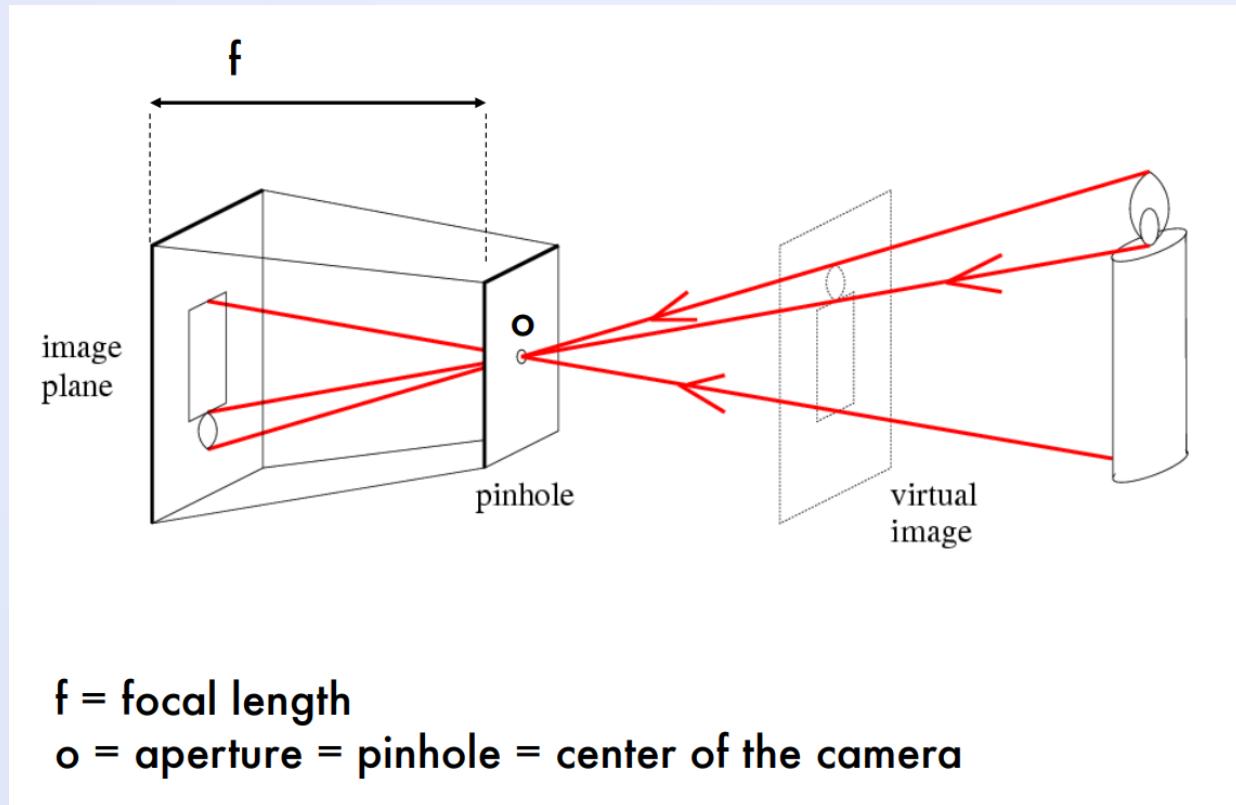


Modélisation de la camera

Pinhole/sténopé



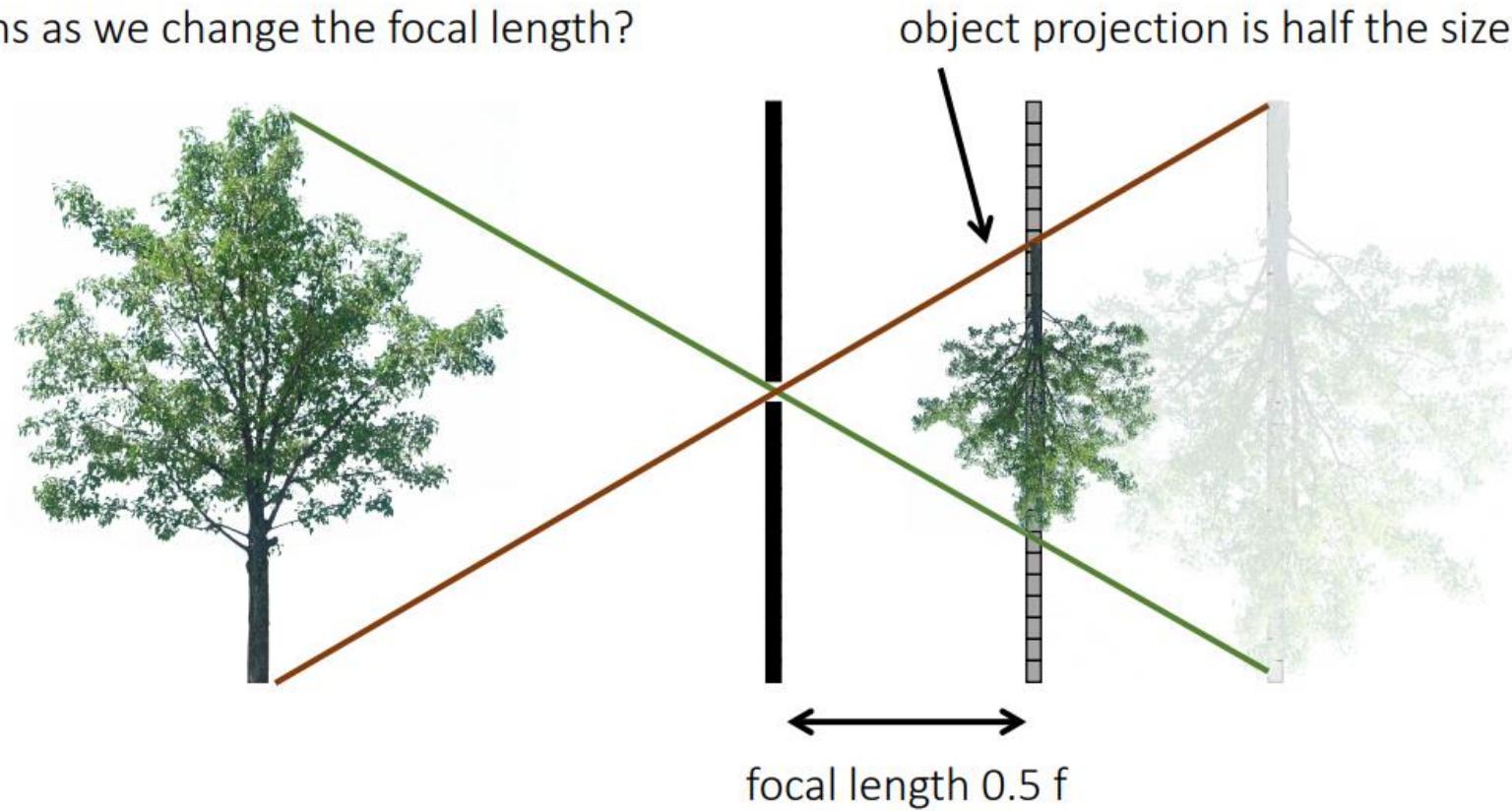
Modèle Sténopé



Distance focale

What happens as we change the focal length?

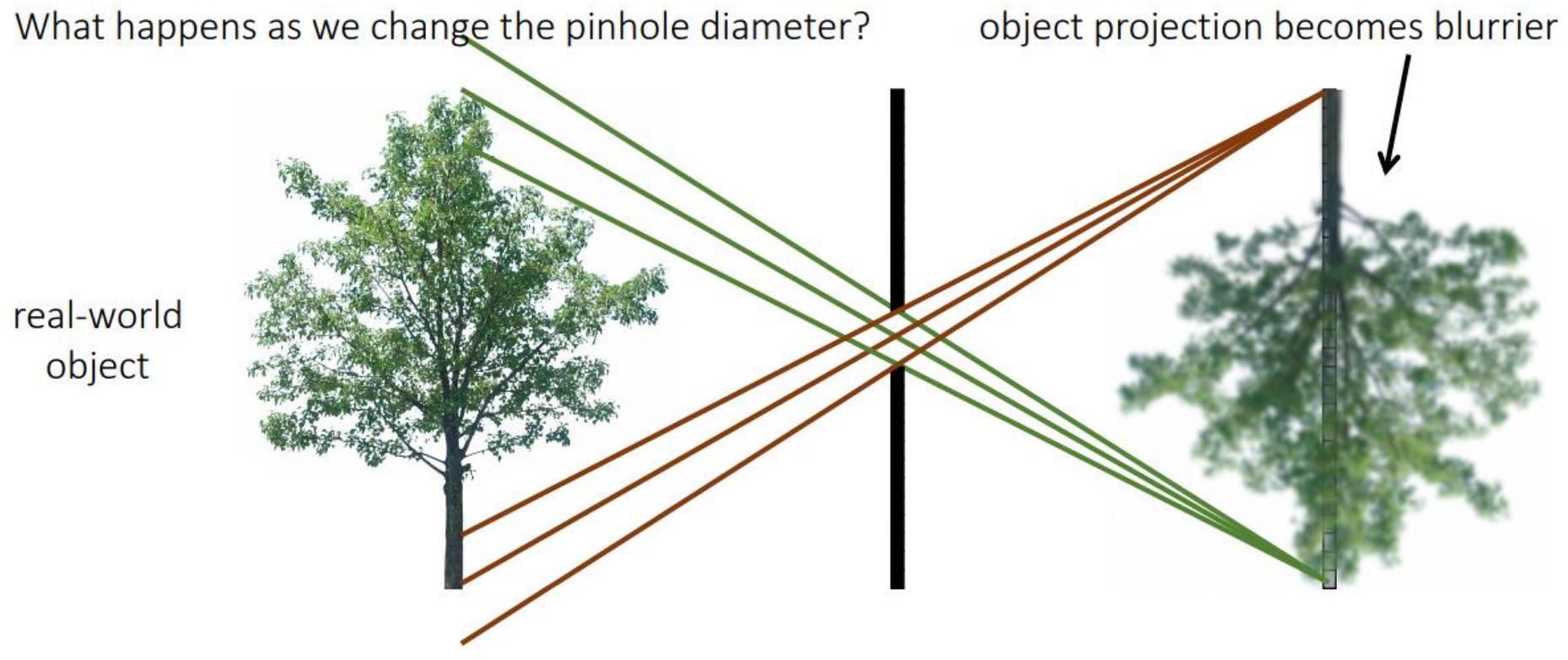
real-world object



focal length $0.5 f$

object projection is half the size

Ouverture



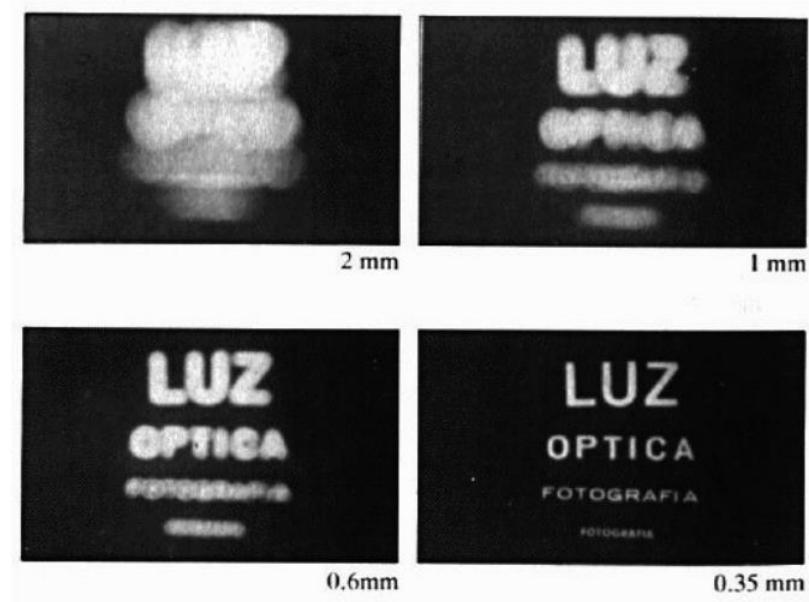
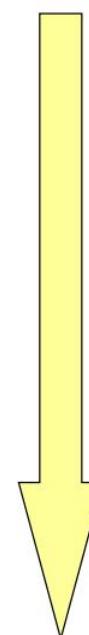
Ouverture

Réduction de la taille de l'ouverture :

- Amélioration de la netteté
- Réduction de la luminosité

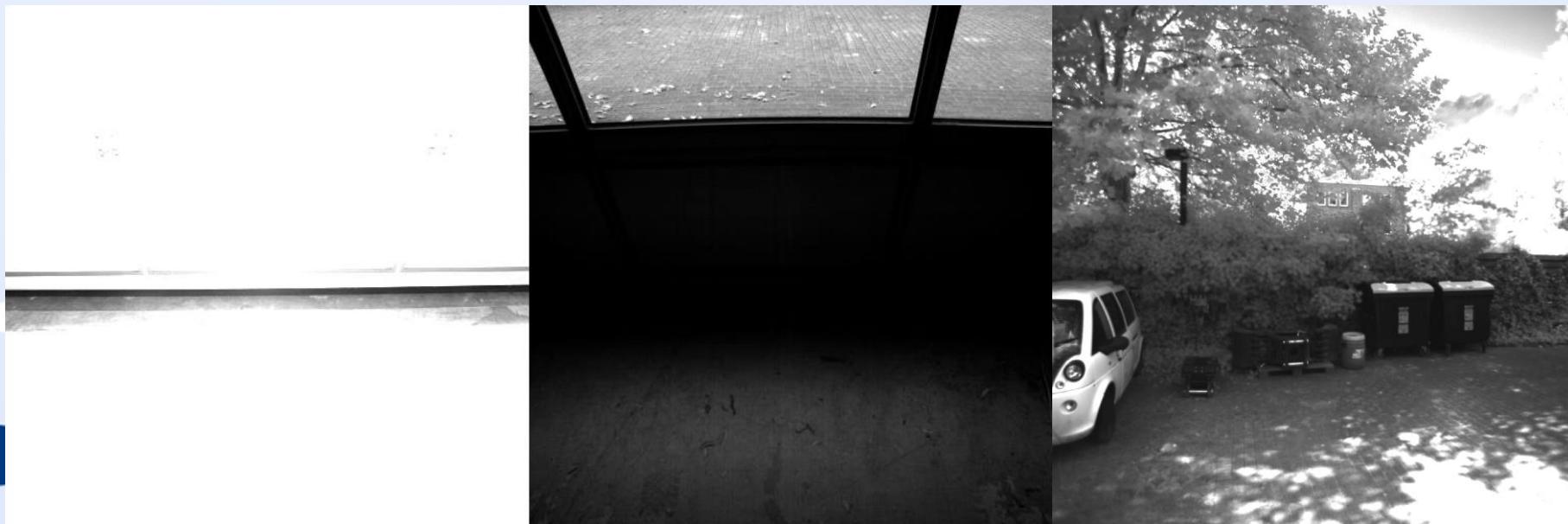


**Shrinking
aperture
size**



Lumière

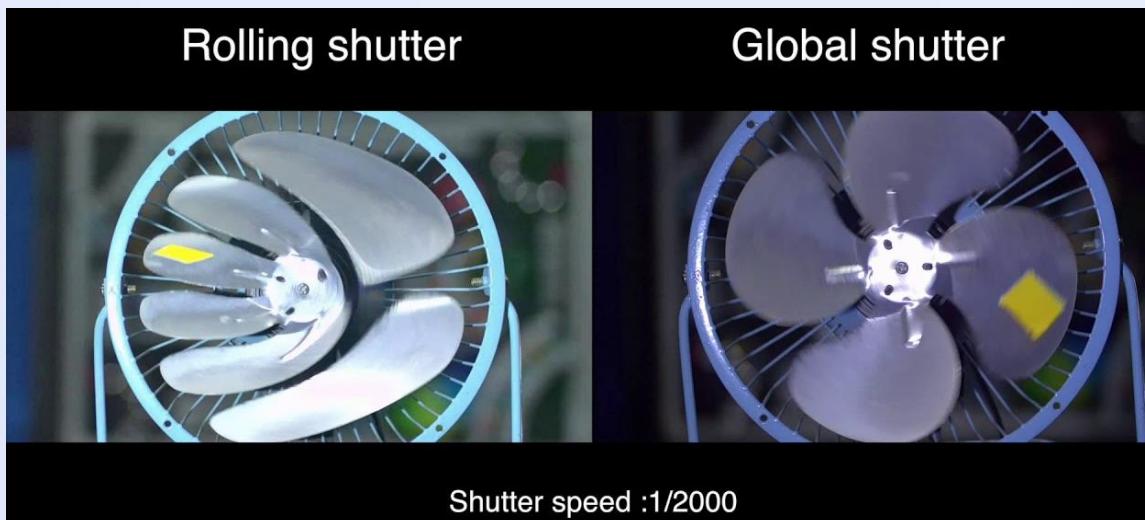
- Sans lumière, pas d'image
- L'illumination de la scène a une influence importante sur le processus d'acquisition
- Contrôler l'illumination est un concept clef
- Globalement possible de contrôler l'illumination dans l'industrie
- Difficile à impossible en milieu extérieur





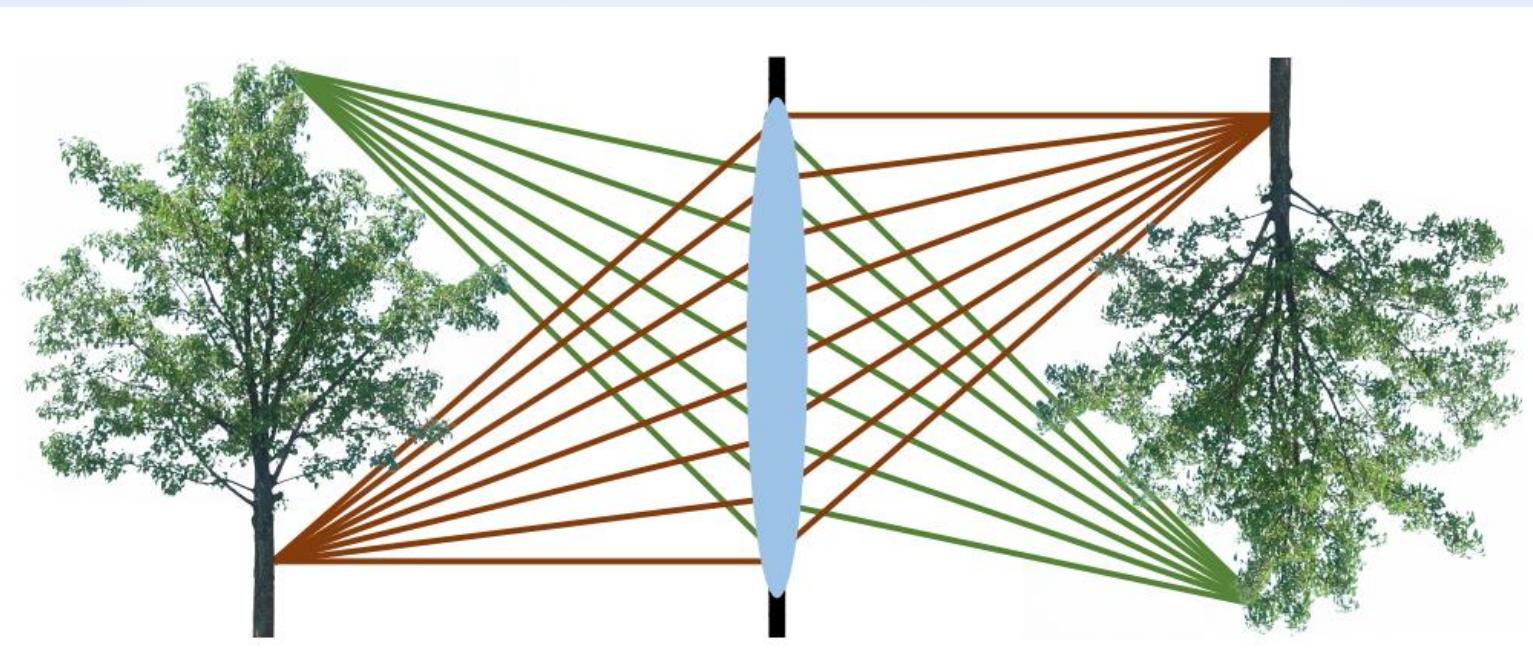
Contrôle de l'acquisition

- 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 »



Lentilles

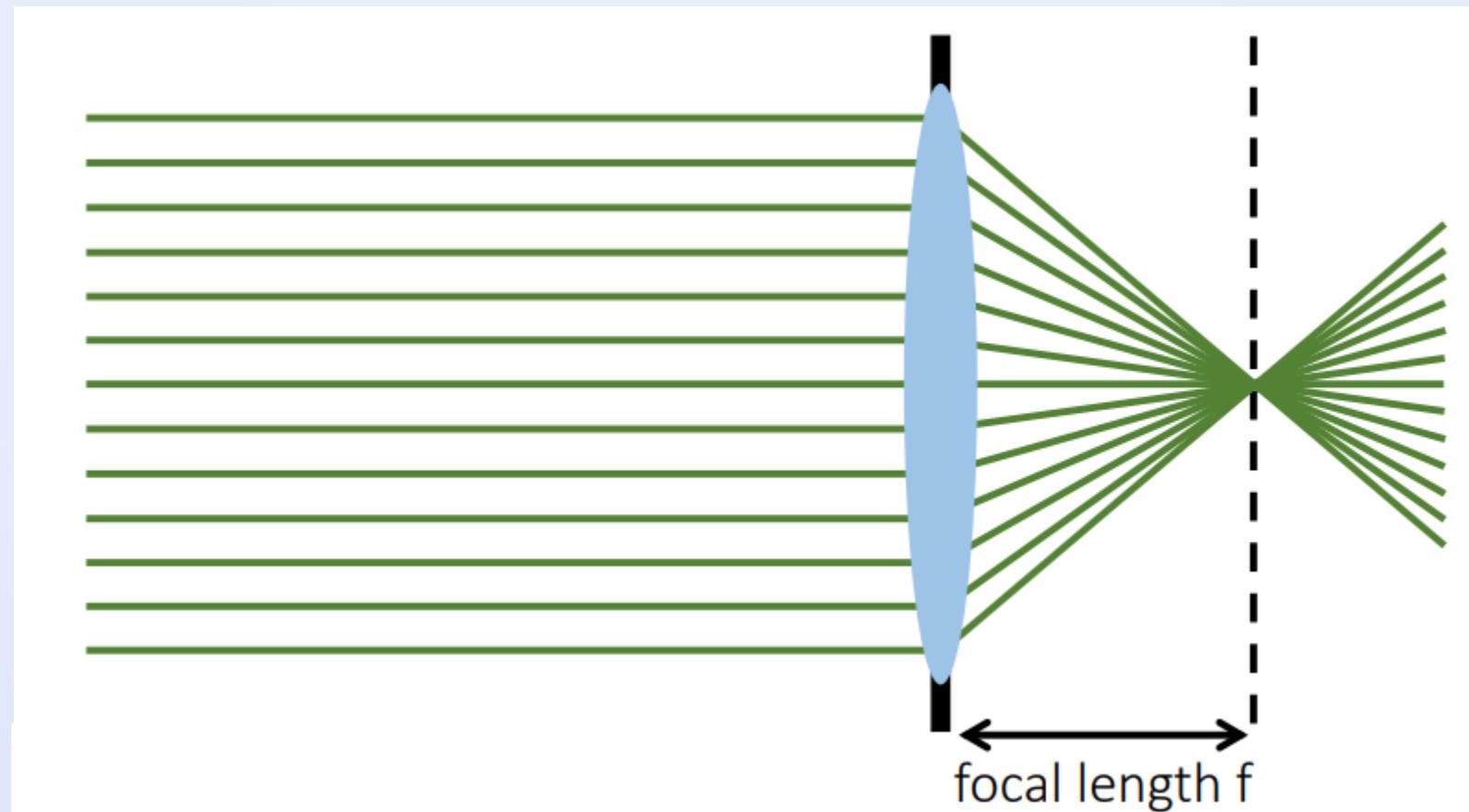
- Utilisation d'une lentille pour concentrer la lumière sur le plan image
 - Amélioration de la luminosité
 - Amélioration de la netteté
- Le choix de la lentille dépend:
 - De la distance de travail
 - Du champ de vision recherché





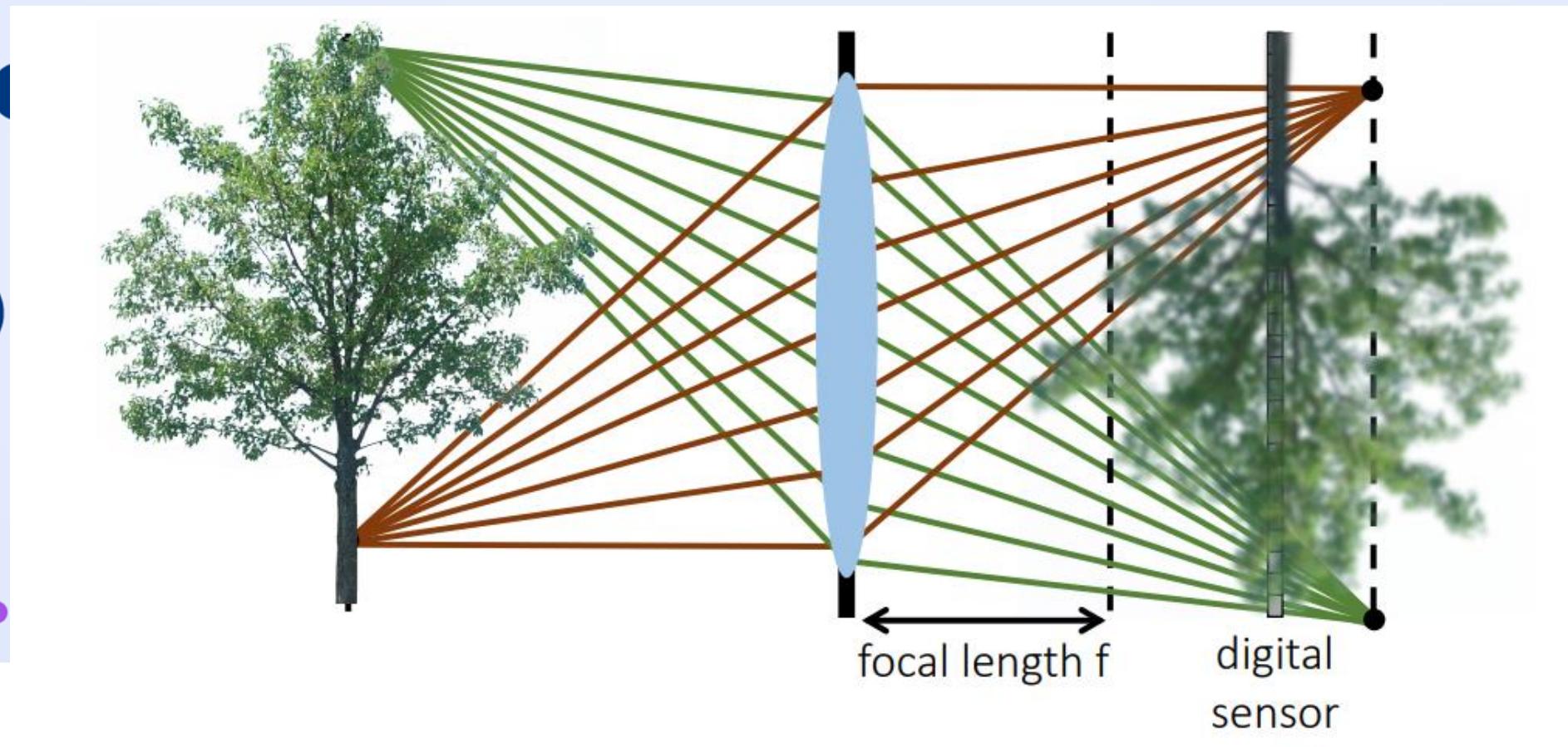
Distance focale

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

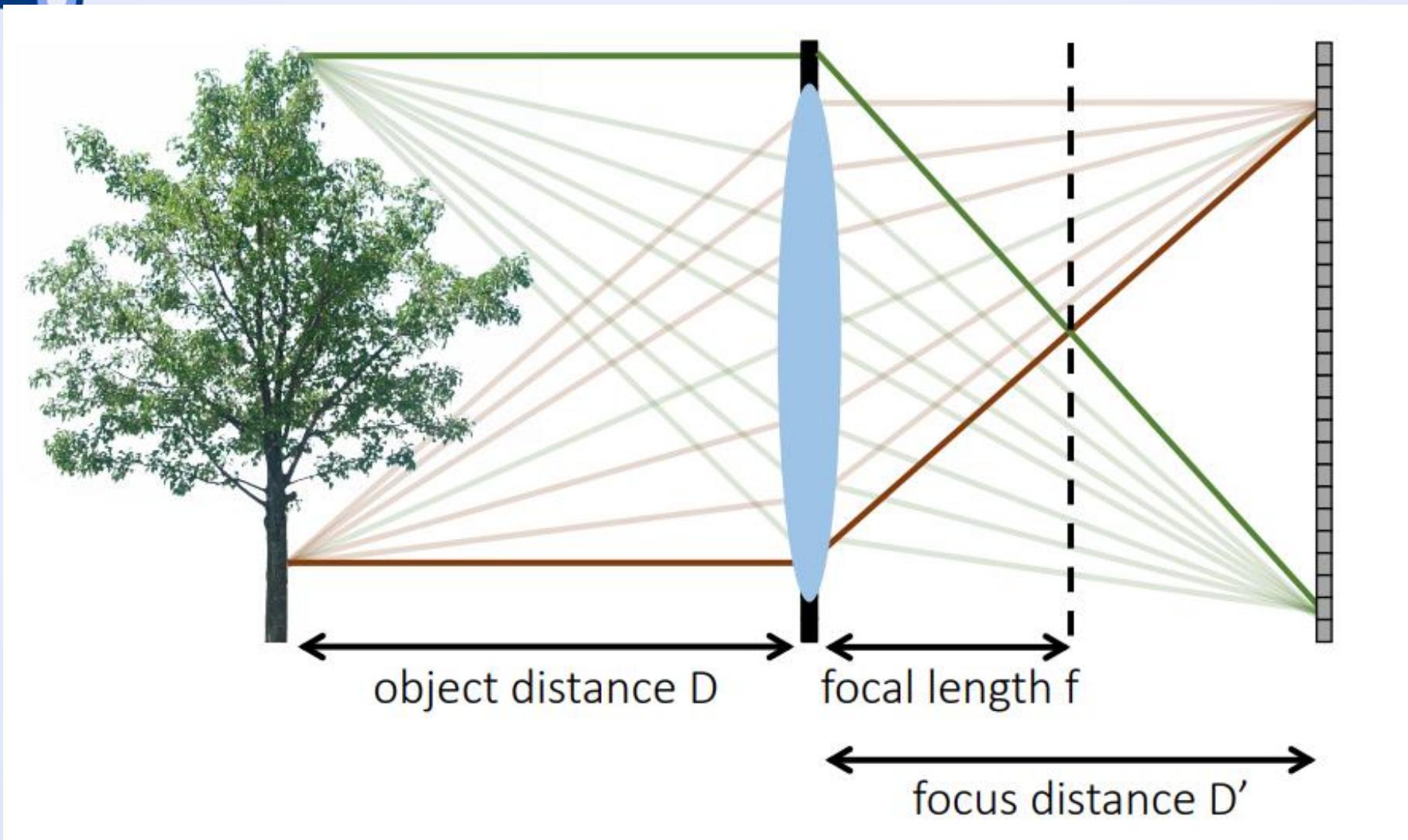


Focus

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

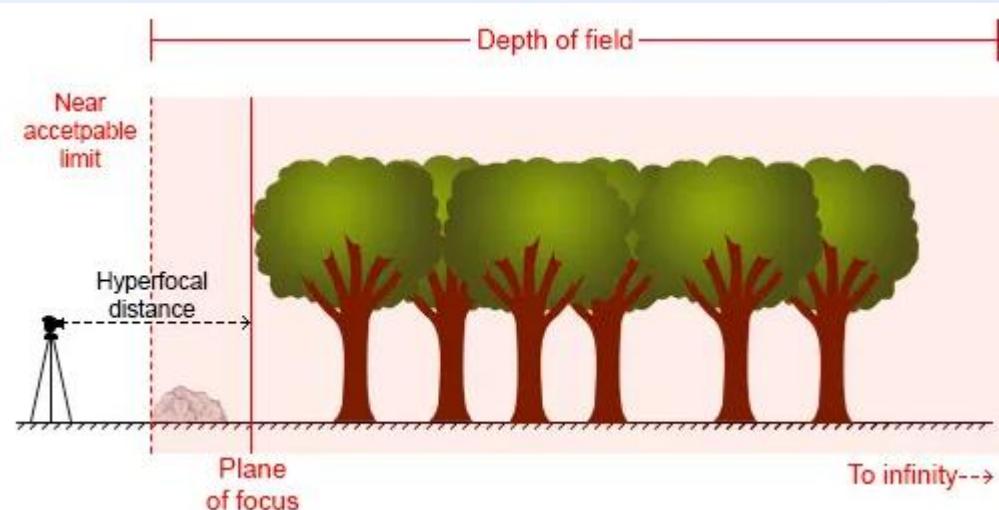
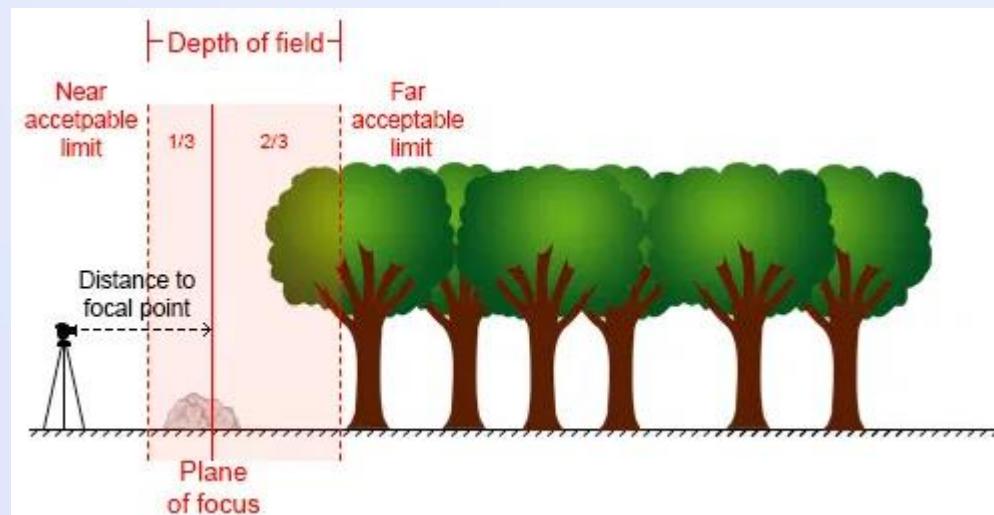


Distance de focus



Hyperfocale

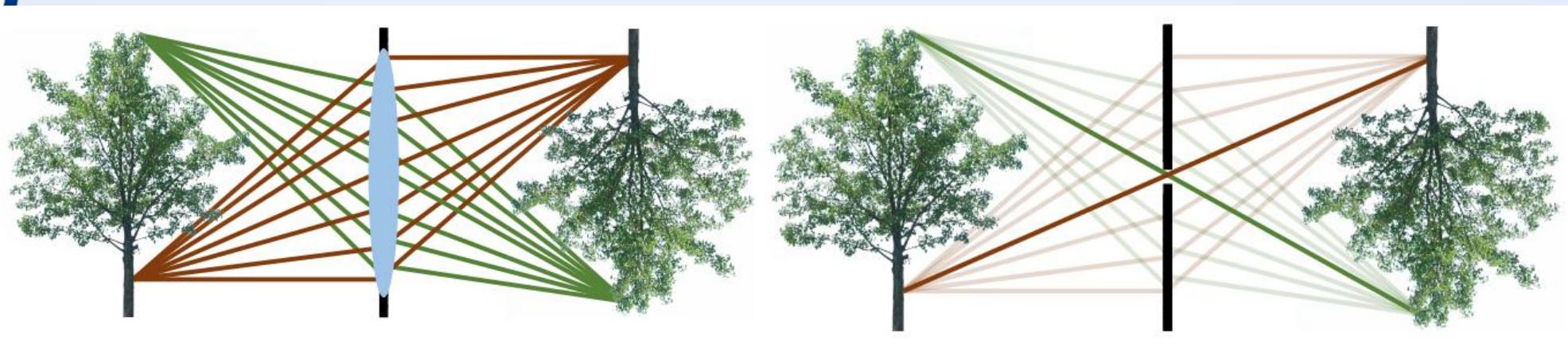
- Distance minimum à laquelle il est possible de faire la mise au point tout en gardant les objets situés à l'infini avec une netteté acceptable.
- Dépends de la distance focale et de l'ouverture



Validité du modèle sténopé

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

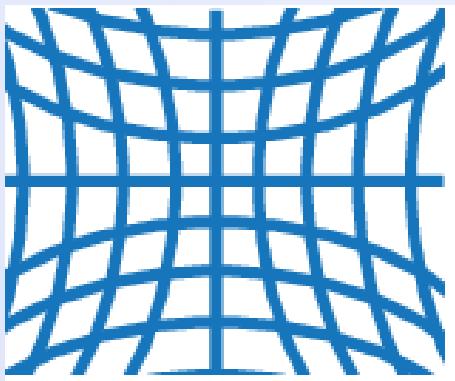




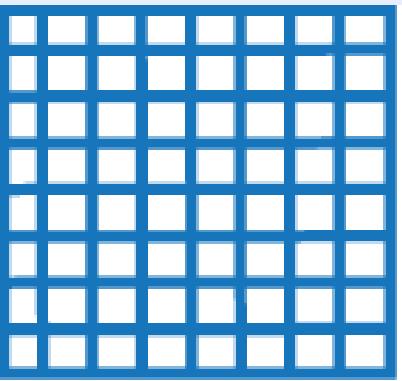
Distortions

Problème des lentilles:

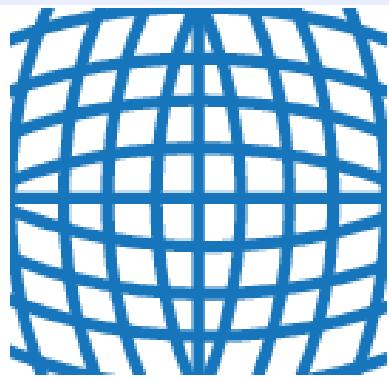
- Distorsion radiale
- La distorsion est plus importante pour les rayons qui passent près du bord de la lentille



Negative radial distortion
"pincushion"



No distortion



Positive radial distortion
"barrel"



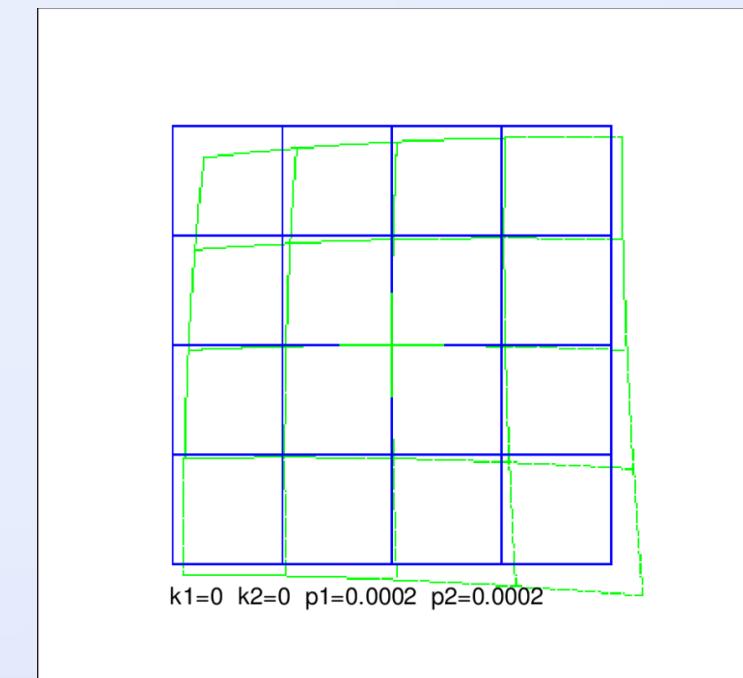
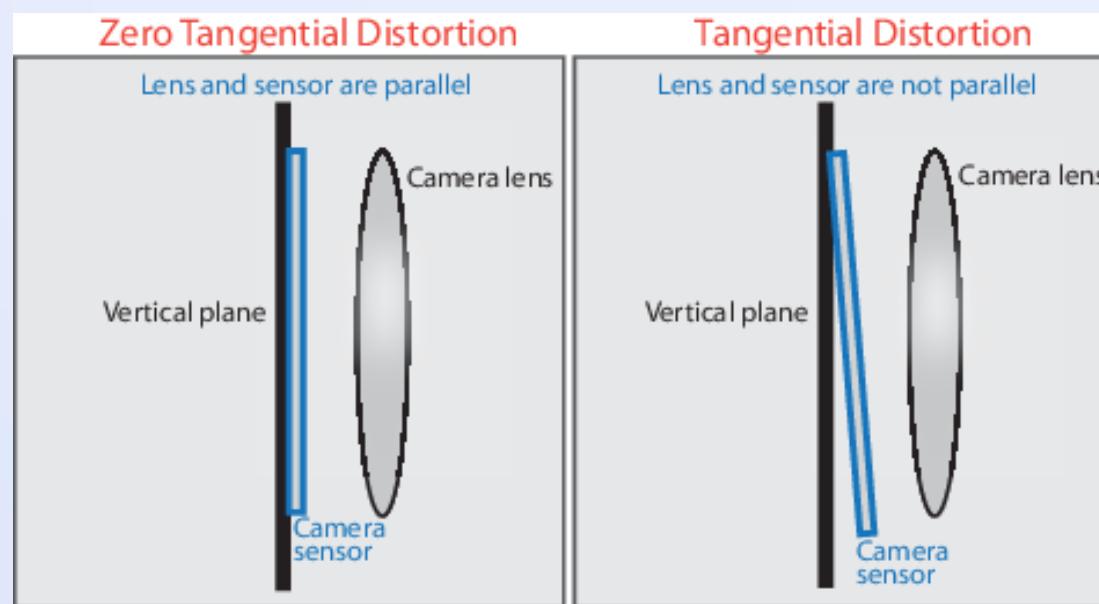
Image magnification decreases with distance from the optical axis



Distortions

Problème des lentilles:

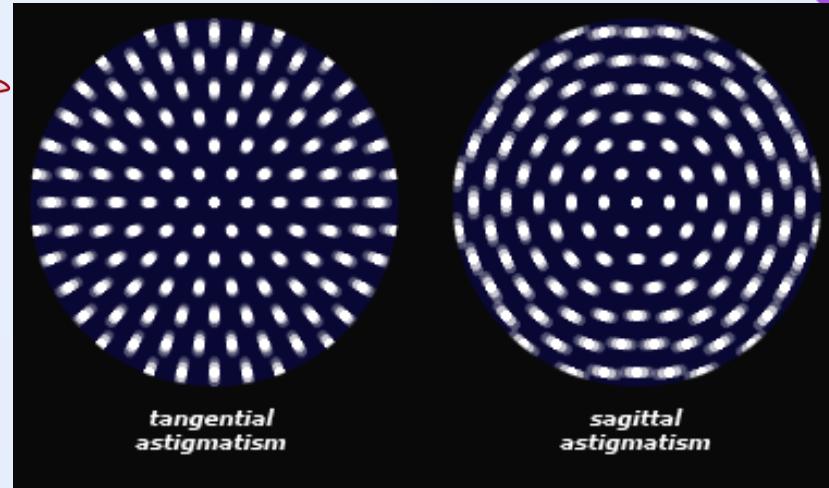
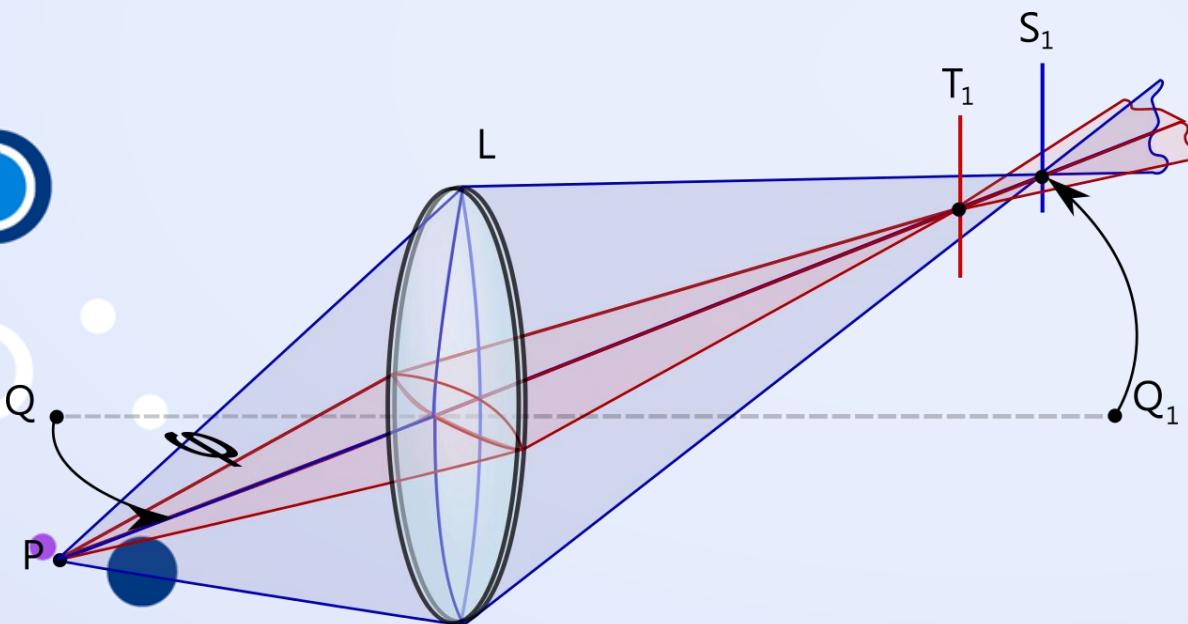
- Distorsion tangentielle
- La distorsion est plus importante pour les rayons qui passent près du bord de la lentille



Distortions

Problème des lentilles:

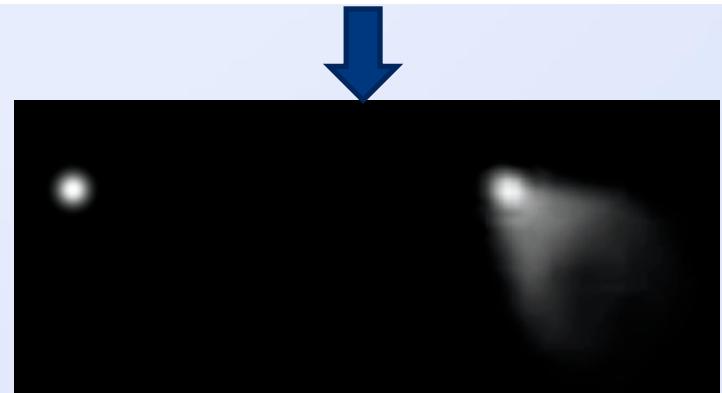
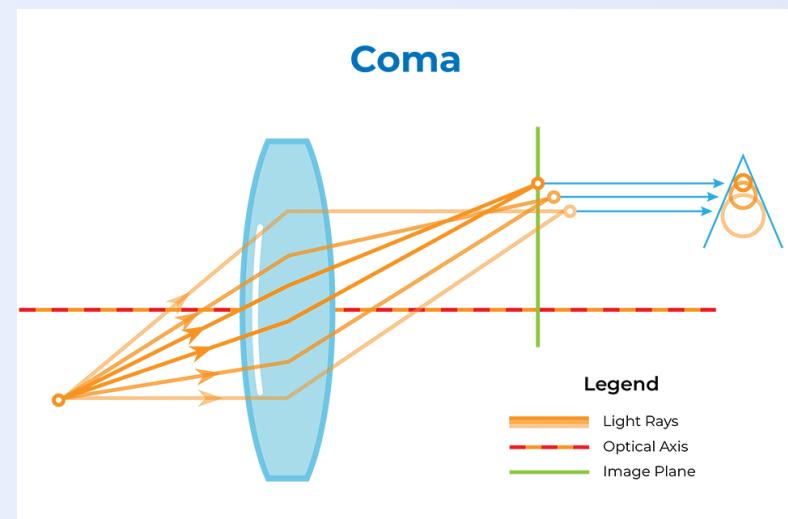
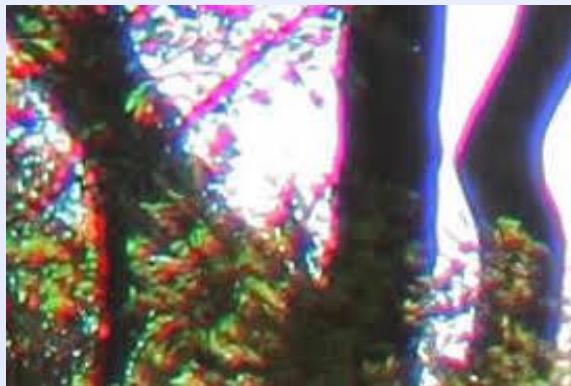
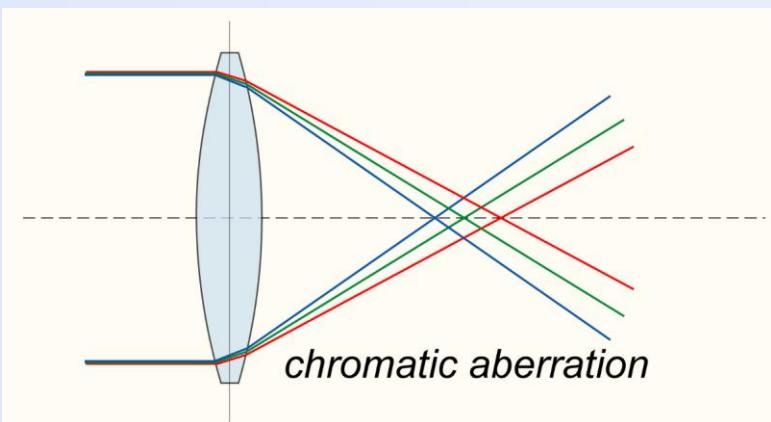
- Astigmatisme
- La distance focale est différente selon l'axe X et l'axe Y car la lentille n'est pas parfaitement circulaire



Aberrations

Problème des lentilles:

- Aberration chromatique
- Aberration comatique



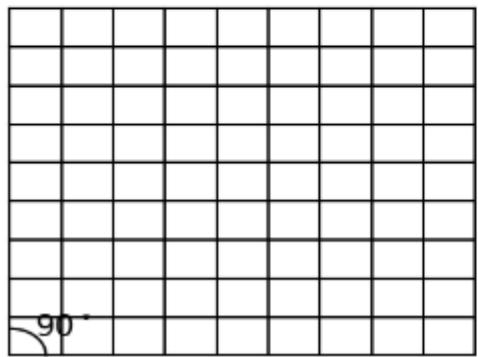


Malformation capteur

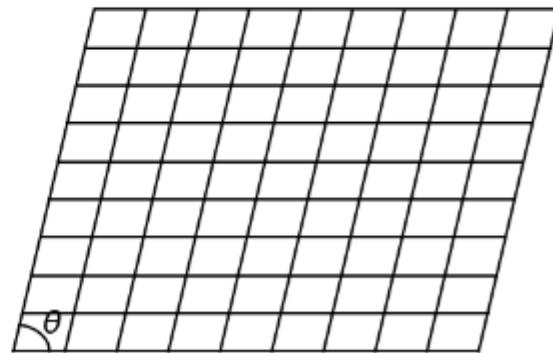
Problème du capteur:

- 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)



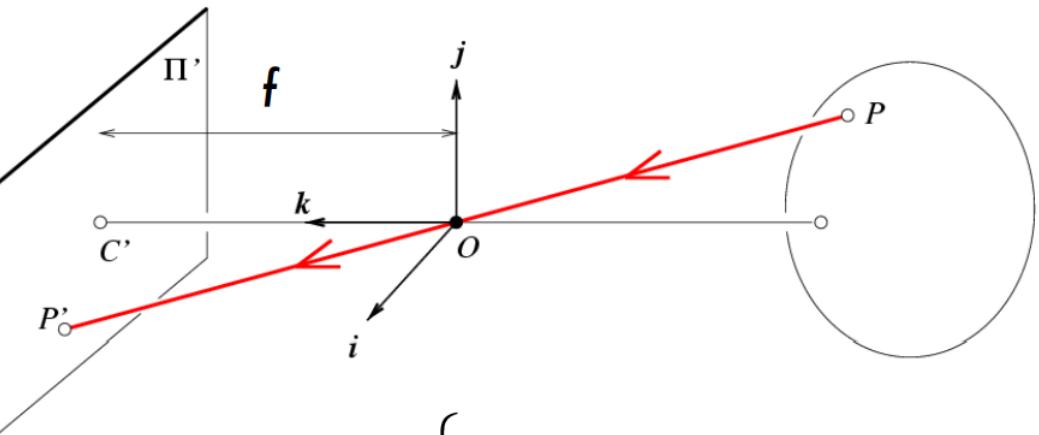
Bruit

Problème du capteur:

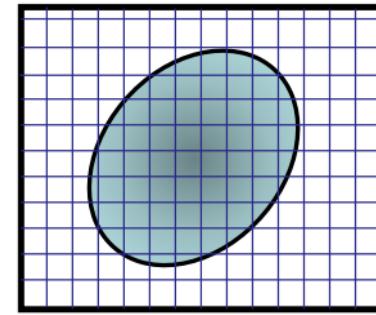
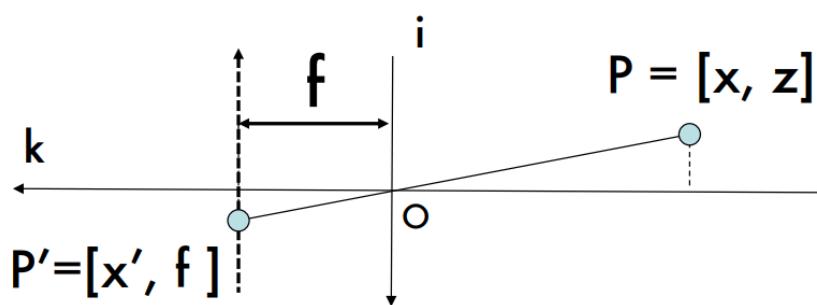
- Bruit lié à l'électronique
- Bruit lié à la discrétisation des mesures
(1 octet par pixel => seulement 255 valeurs possibles)



Géométrie optique



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



Pixels, bottom-left coordinate systems

Digital image

f = focal length
 \circ = center of the camera

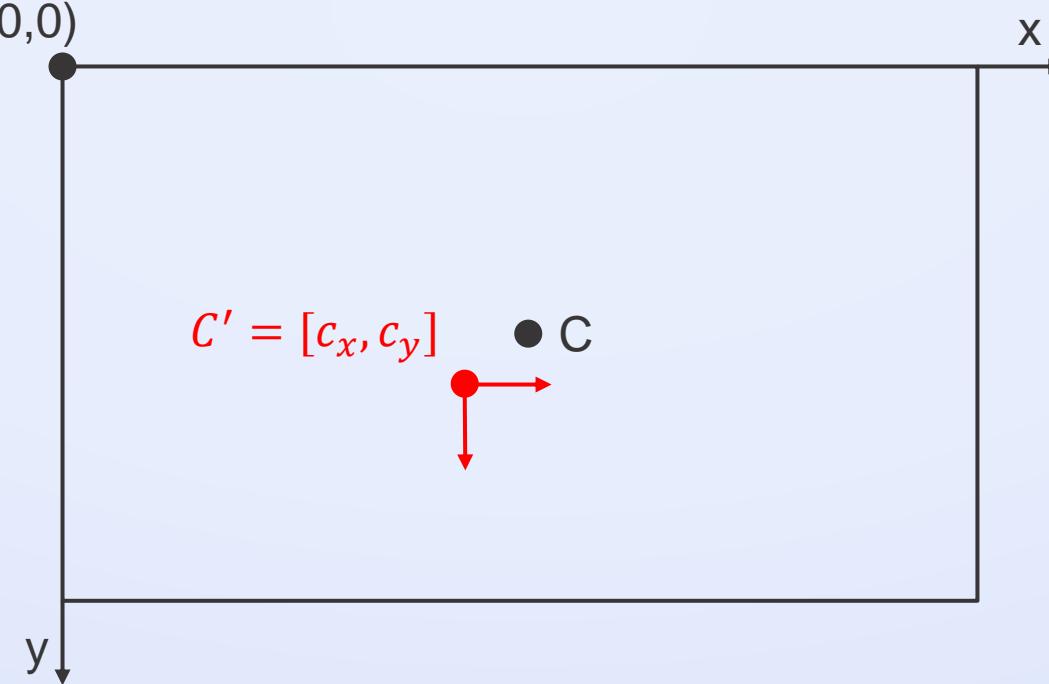
$$\frac{x'}{f} = \frac{x}{z}$$

Système d'équations

- 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
 - $u = \frac{f_x}{d_x} \cdot \frac{X}{Z} = \alpha_x \cdot \frac{X}{Z}$
 - $v = \frac{f_y}{d_y} \cdot \frac{Y}{Z} = \alpha_y \cdot \frac{Y}{Z}$

Centre optique

- 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





Paramètres intrinsèques

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

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

Matrice intrinsèque



Paramètres intrinsèques

- 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

- $\exists \lambda \text{ tq. } \lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$

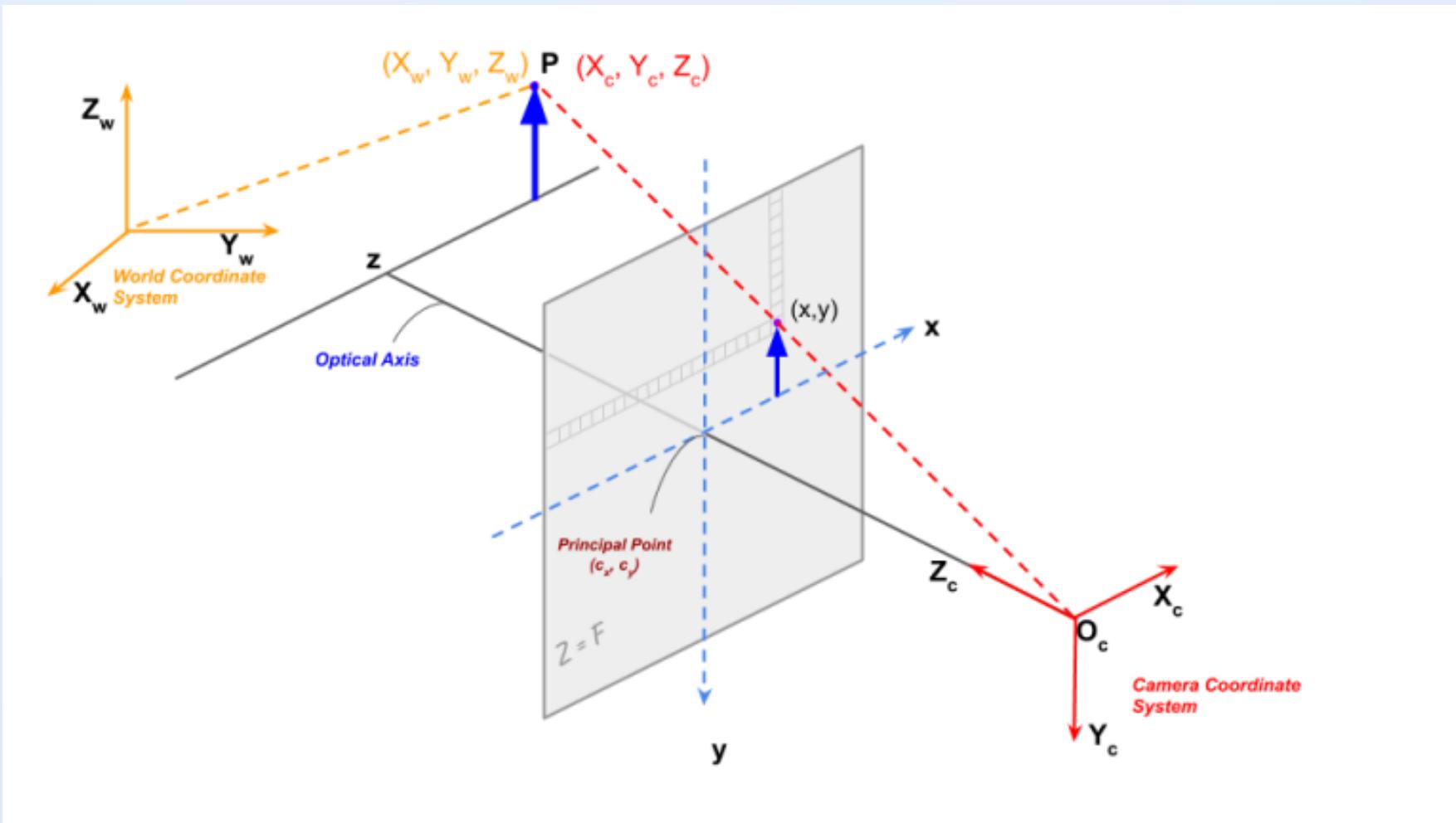
- Preuve:

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

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

Paramètres extrinsèques

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





Du monde à la caméra

Il faut convertir du système 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:

- $\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = [R|t] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Rightarrow \overline{X}_c = [R|t] \cdot \bar{X}$
- $[R|t]$ 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



Du monde à l'image

- 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:

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

- Avec P la matrice caméra 3x4

- D'où: $\bar{x} \sim P\bar{X}$



Estimation de la matrice caméra

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:

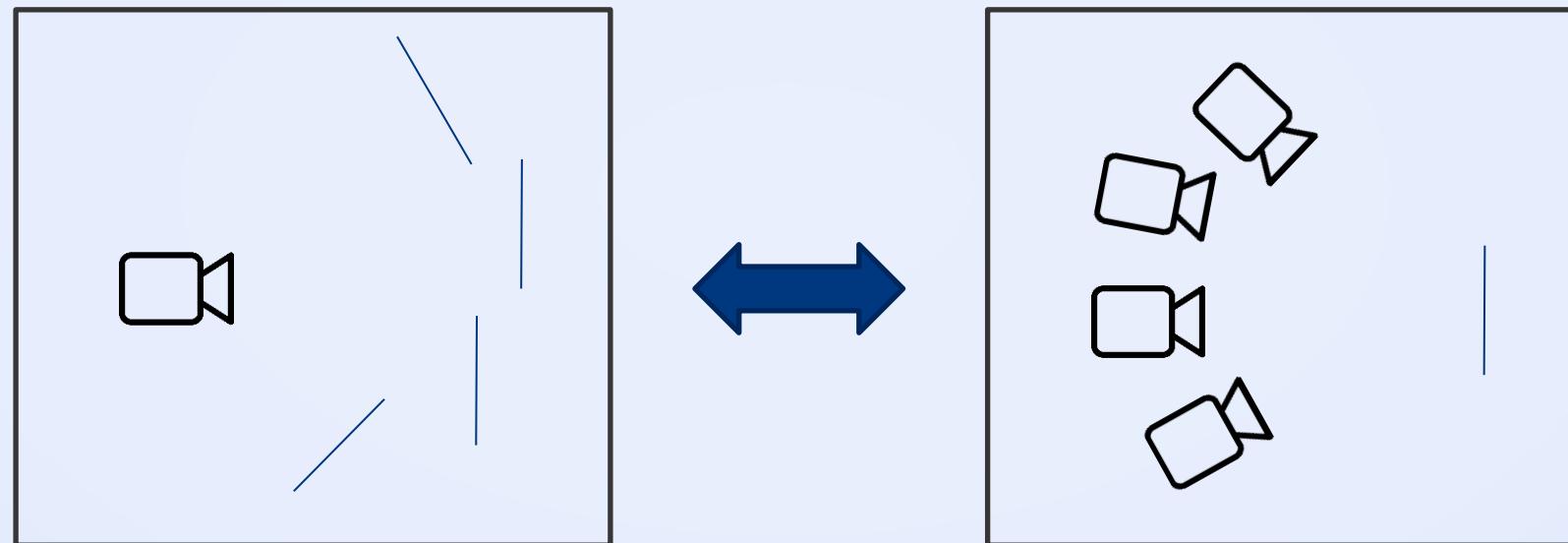
- On estime P à partir de paires \bar{x} et \bar{X} connues
- $\bar{x} \sim P \bar{X}$ donne deux équations indépendantes à 12 inconnues pour chaque point
- On résous $A p = 0$
 - Avec p le vecteur de 12 inconnues
 - A la matrice $2n \times 12$ représentant les équations indépendantes
 - n le nombre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



Estimation de la matrice intrinsèque

Plane-based calibration:

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



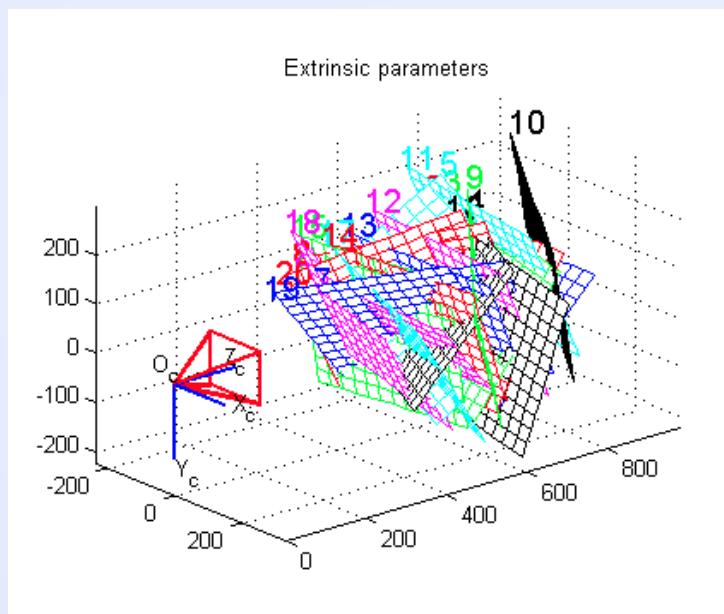
- Cela permet de fixer $Z=0$ pour tous les points du plan observé



Estimation de la matrice intrinsèque

Plane-based calibration:

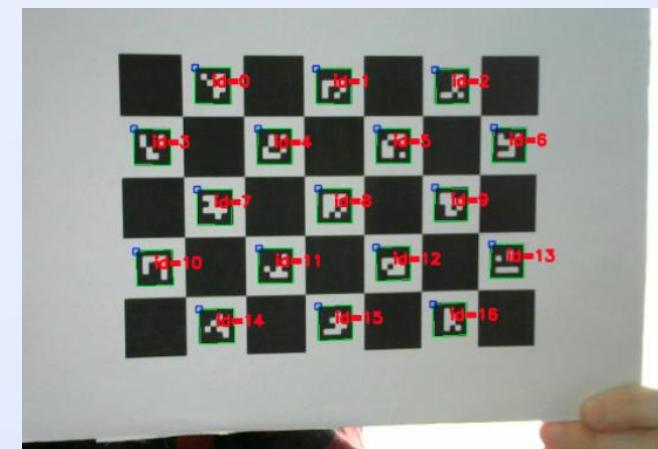
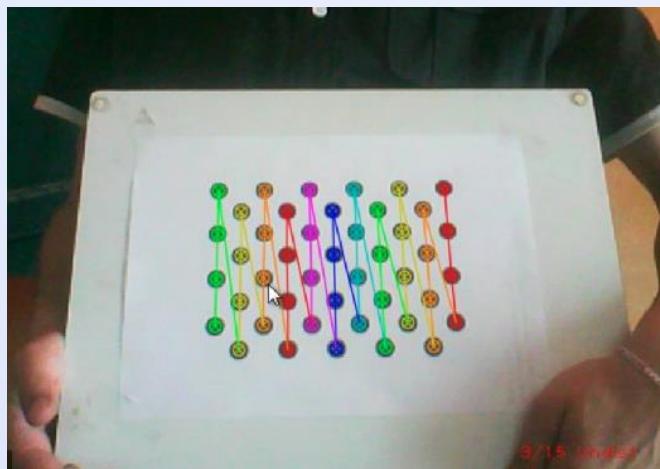
- On réalise l'acquisition de multiples images d'une surface plane (e.g. damier)



Estimation de la matrice intrinsèque

On a donc:

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



Paramètres de distorsion

- Il est aussi possible d'estimer des paramètres caractérisant les distorsions radiales et tangentialles.
- 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 tangentialles:
 - $x_{distorted} = x + 2p_1xy + p_2(r^2 + 2x^2)$
 - $y_{distorted} = y + p_1(r^2 + 2y^2) + 2p_2xy$



Modèle complet

Modèle de distorsions de Brown-Conrady:

$$x_{distorted} = x \cdot \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + 2p_1 xy + p_2(r^2 + 2x^2)$$

$$y_{distorted} = y \cdot \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + p_1(r^2 + 2y^2) + 2p_2 xy$$

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 $x = \frac{X_c}{Z_c}, y = \frac{Y_c}{Z_c}$

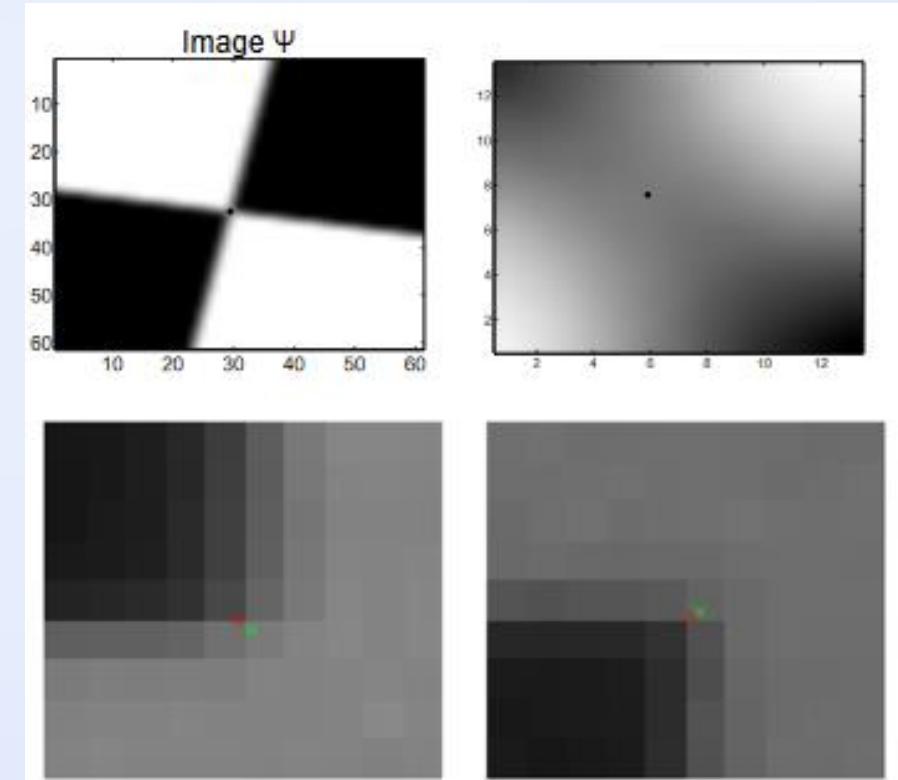
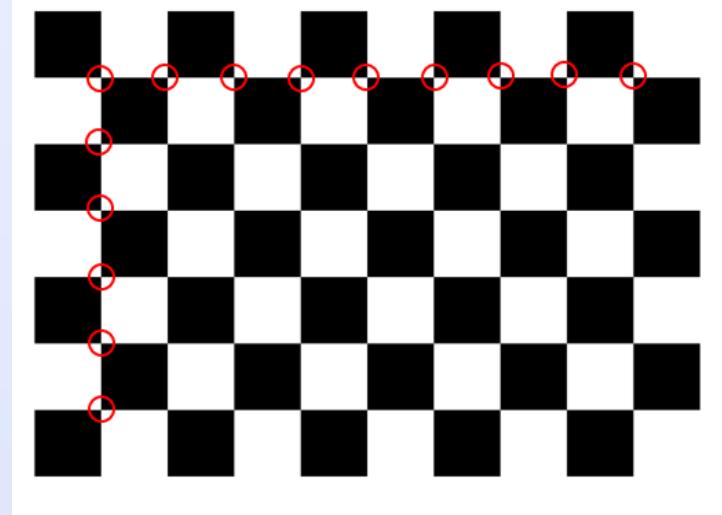
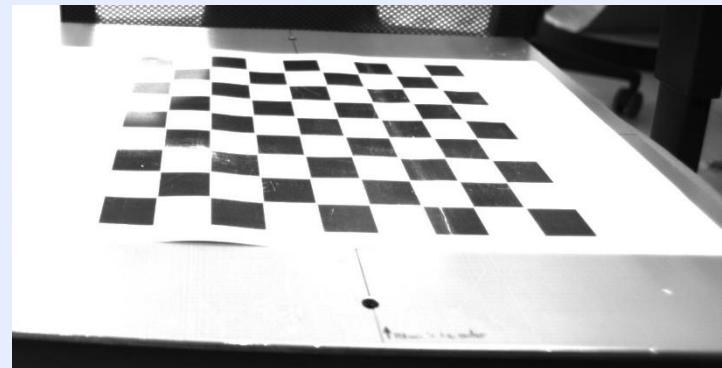
Modèle complet:

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



Comment réaliser une bonne calibration?

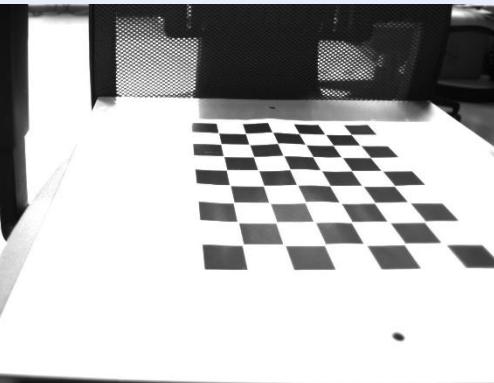
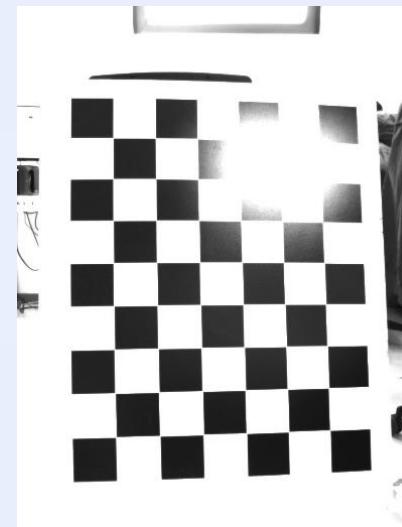
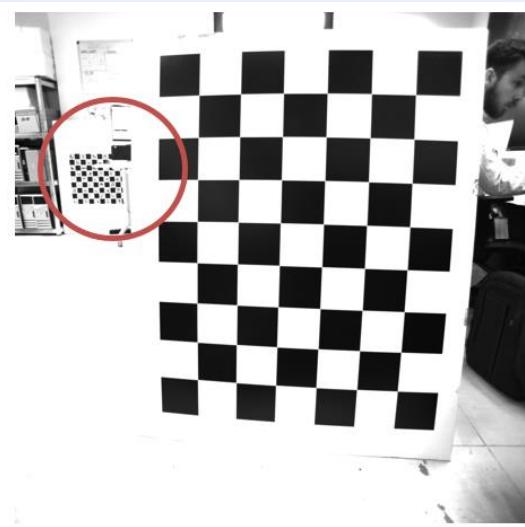
- La cible doit être parfaitement plane
- Les motifs de type cercles asymétriques donnent de meilleurs résultats
- La distance entre les points doit être mesurée précisément
- Il faut correctement définir le nombre de lignes/colonnes de la cible





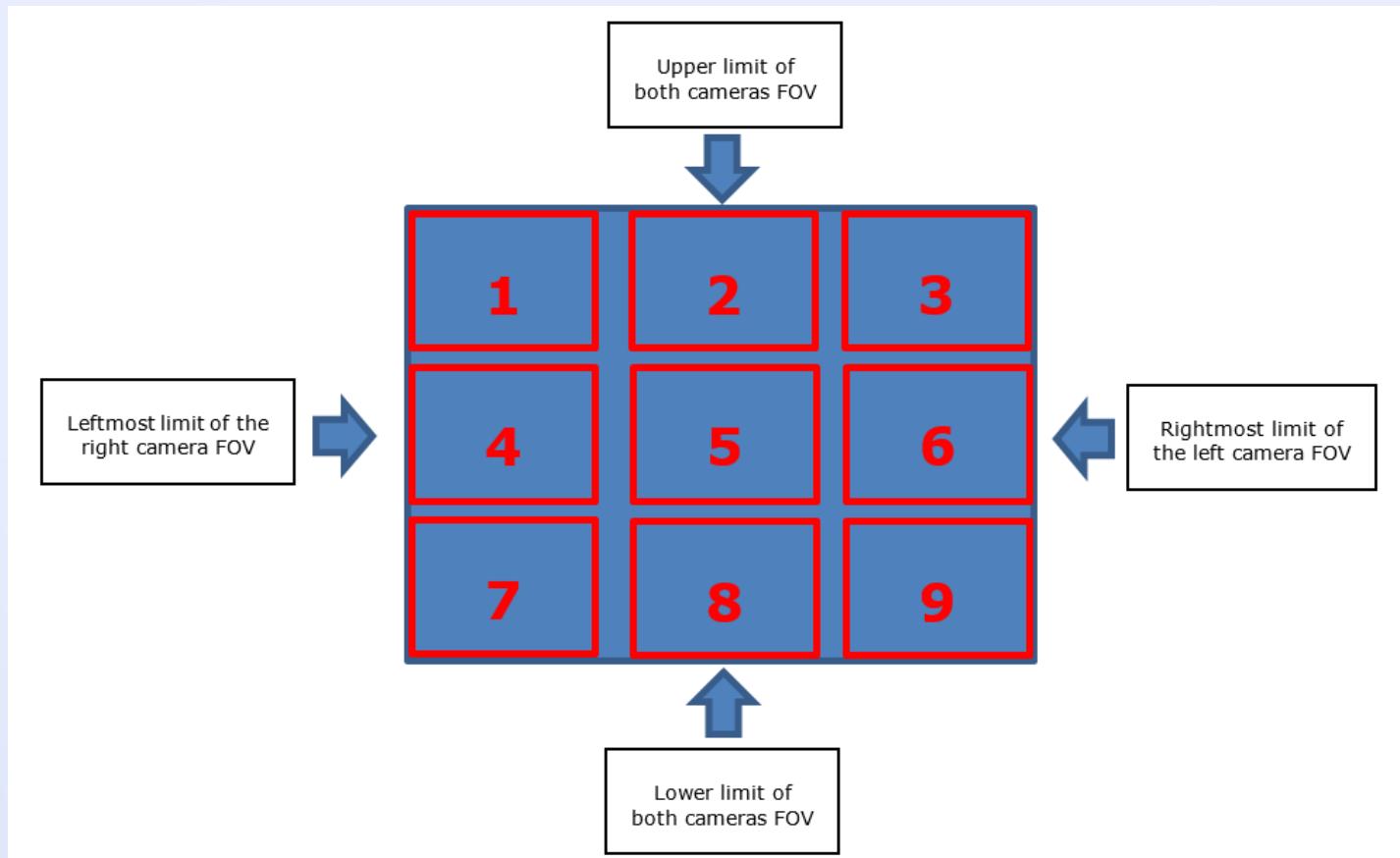
Contrôler l'environnement de calibration

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

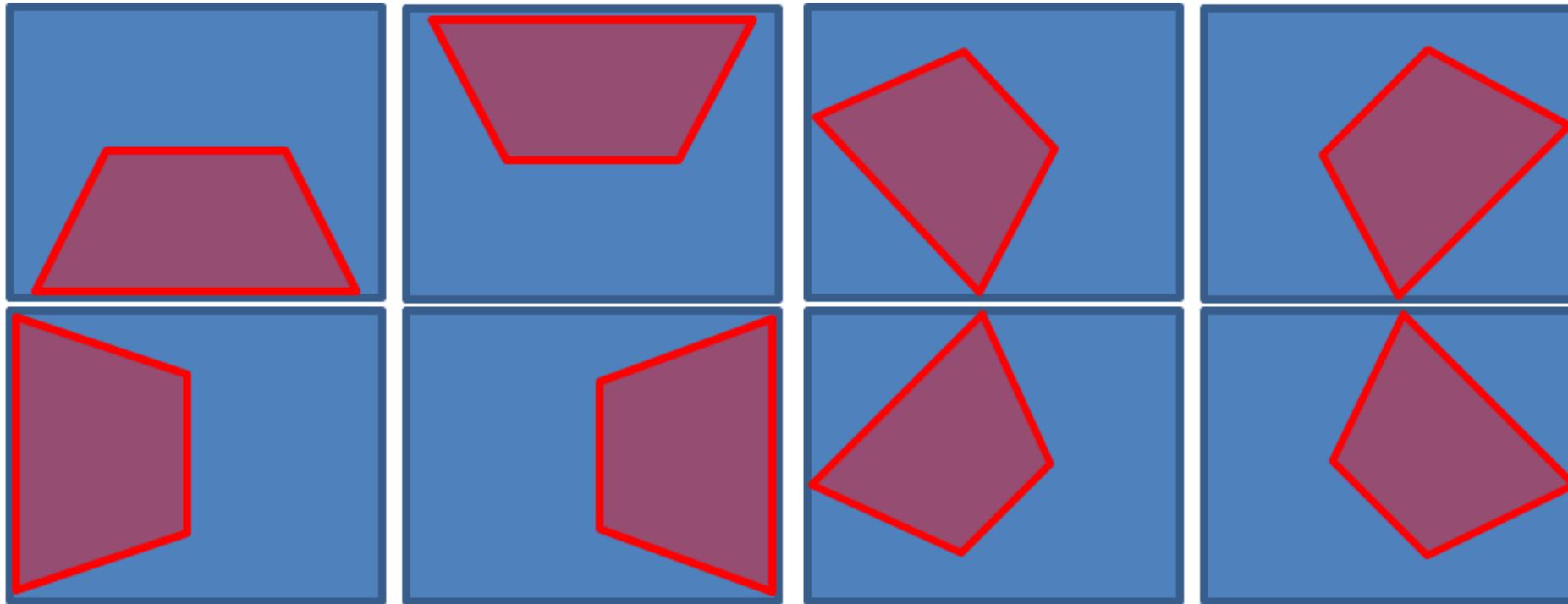


Procédure de calibration

- Mettre la caméra dans une position fixe
- Acquérir (au moins) 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



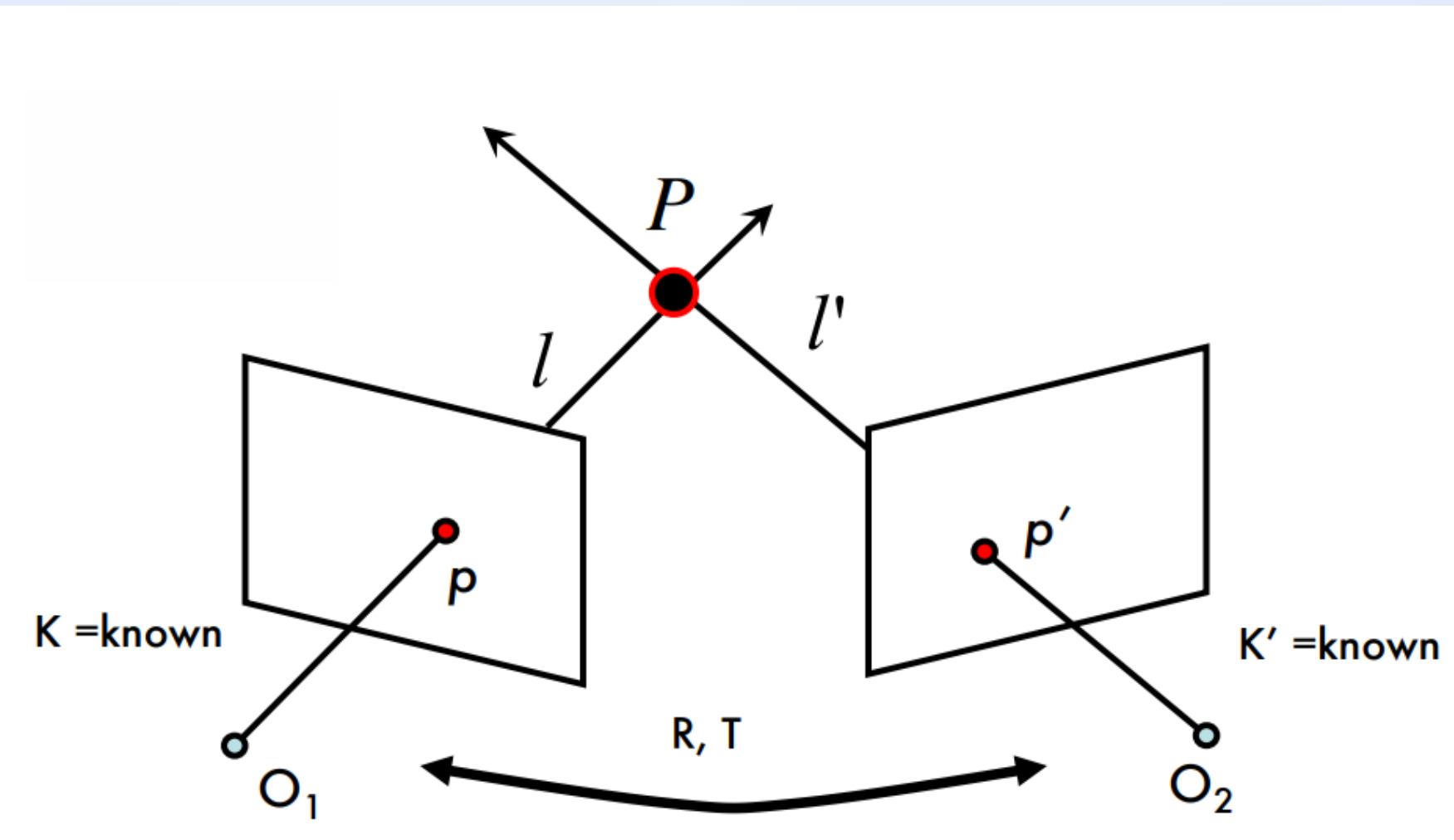
- Acquérir 8 images avec des inclinaisons différentes de la cible
- Acquérir 5 à 10 images supplémentaire avec des angles « aléatoires »
- Essayer d'avoir une distribution homogène des points dans le plan image



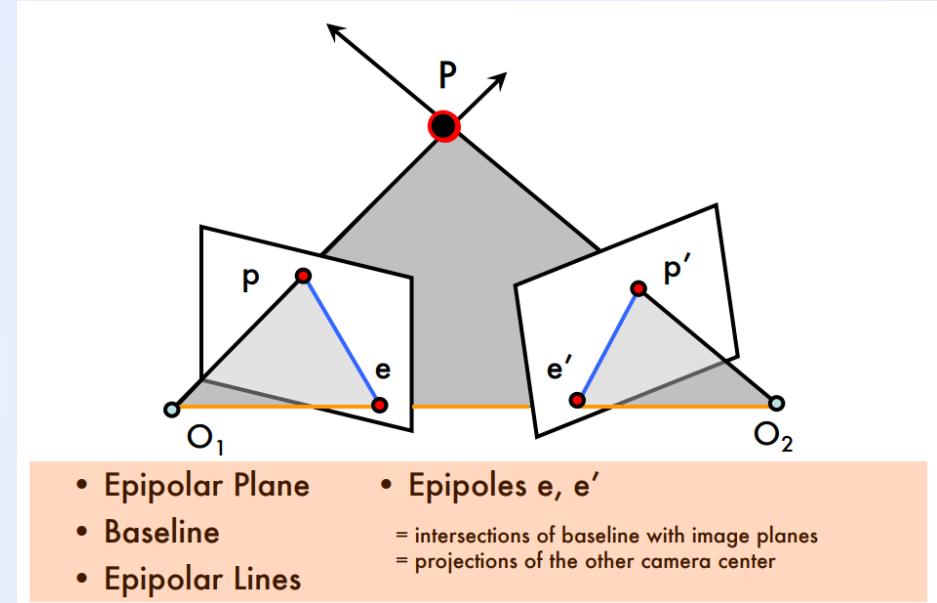
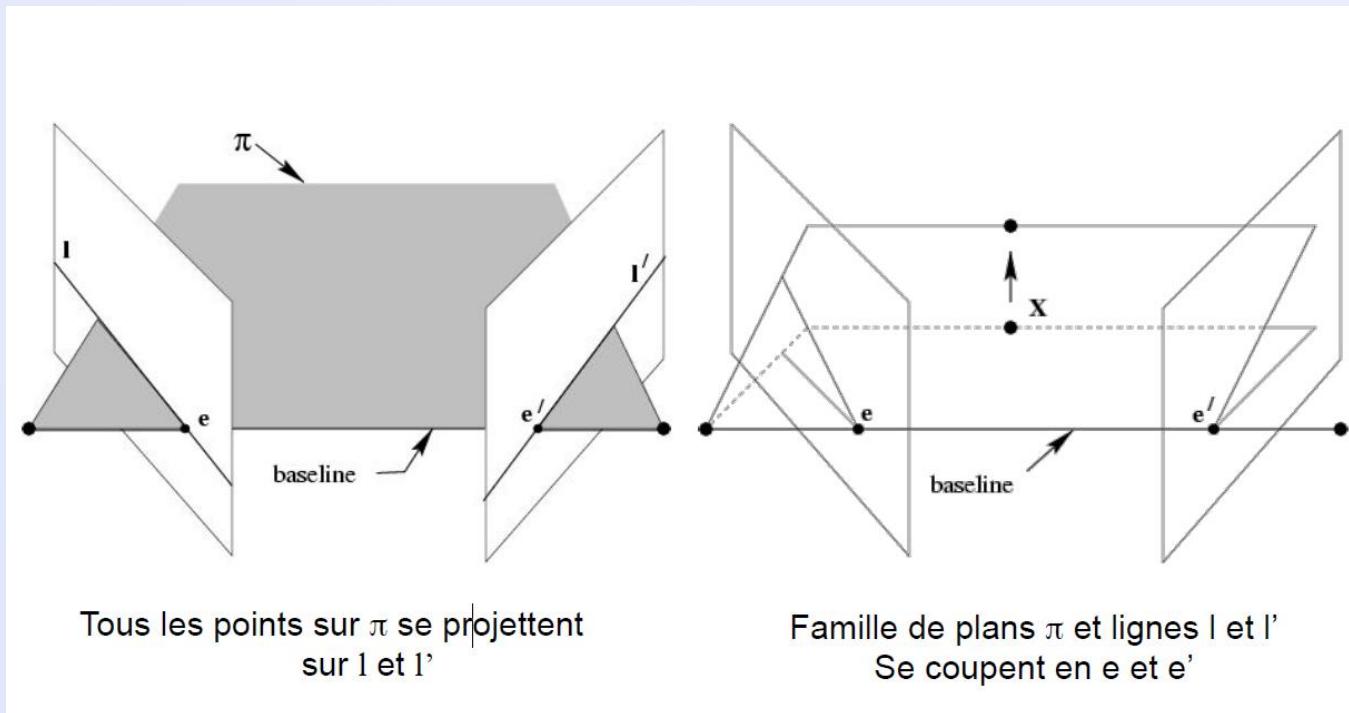
Stéréovision



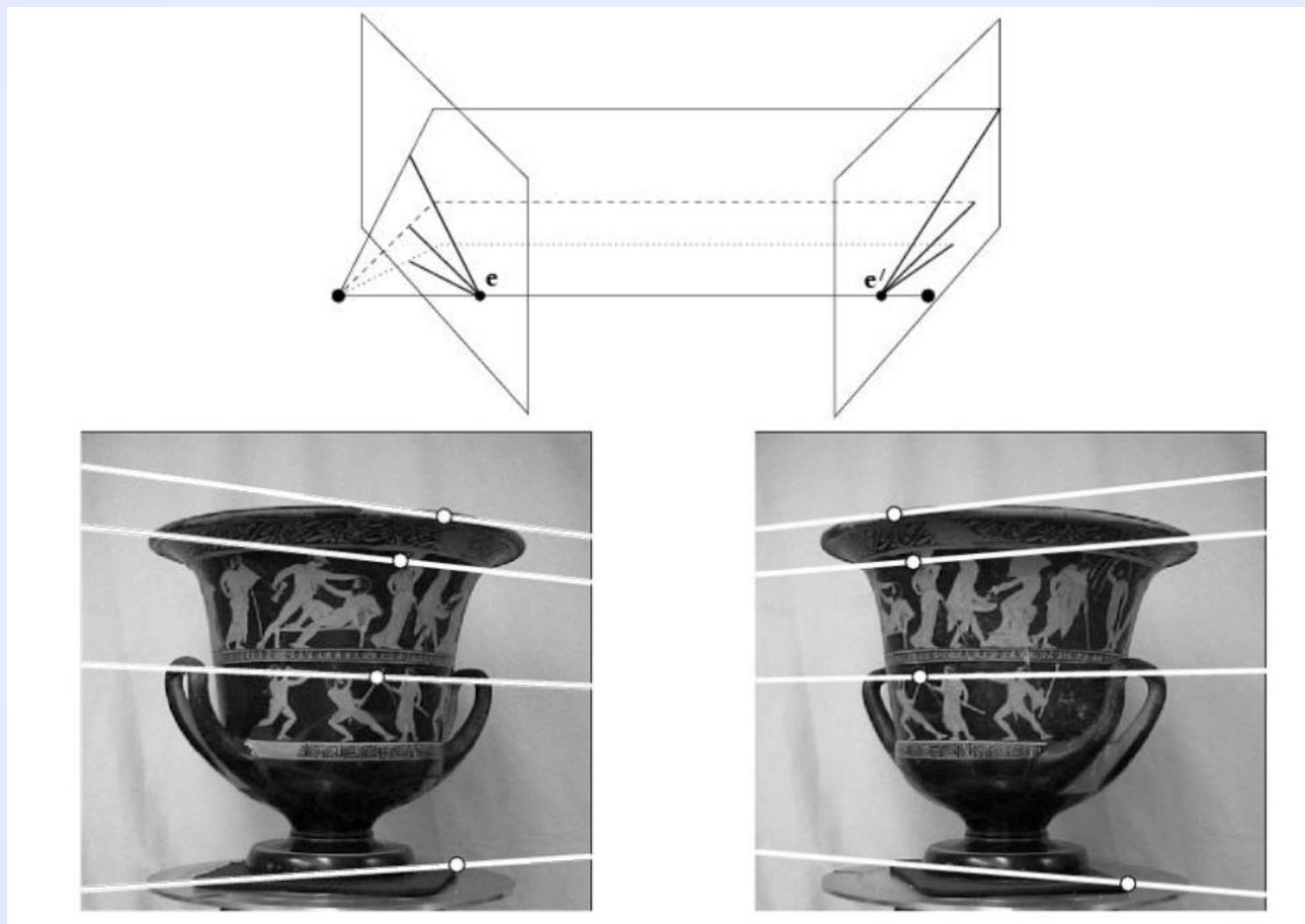
Triangulation



Géométrie épipolaire

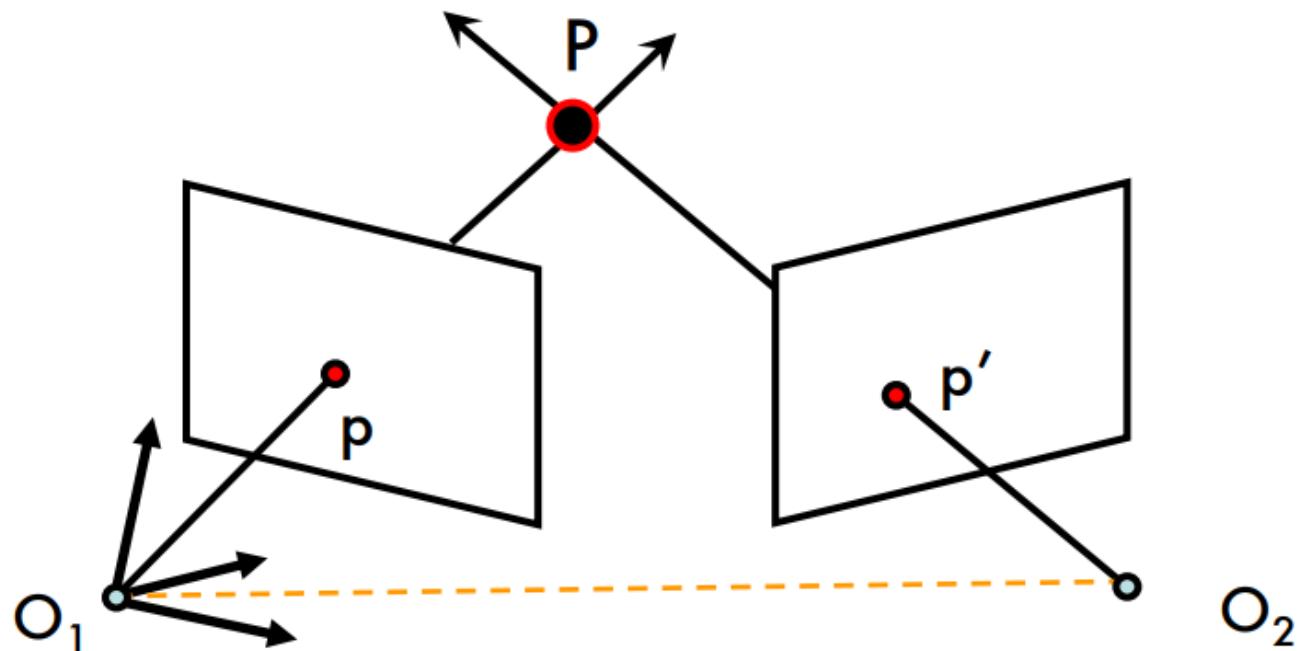


Lignes épipolaire





Matrice essentielle E



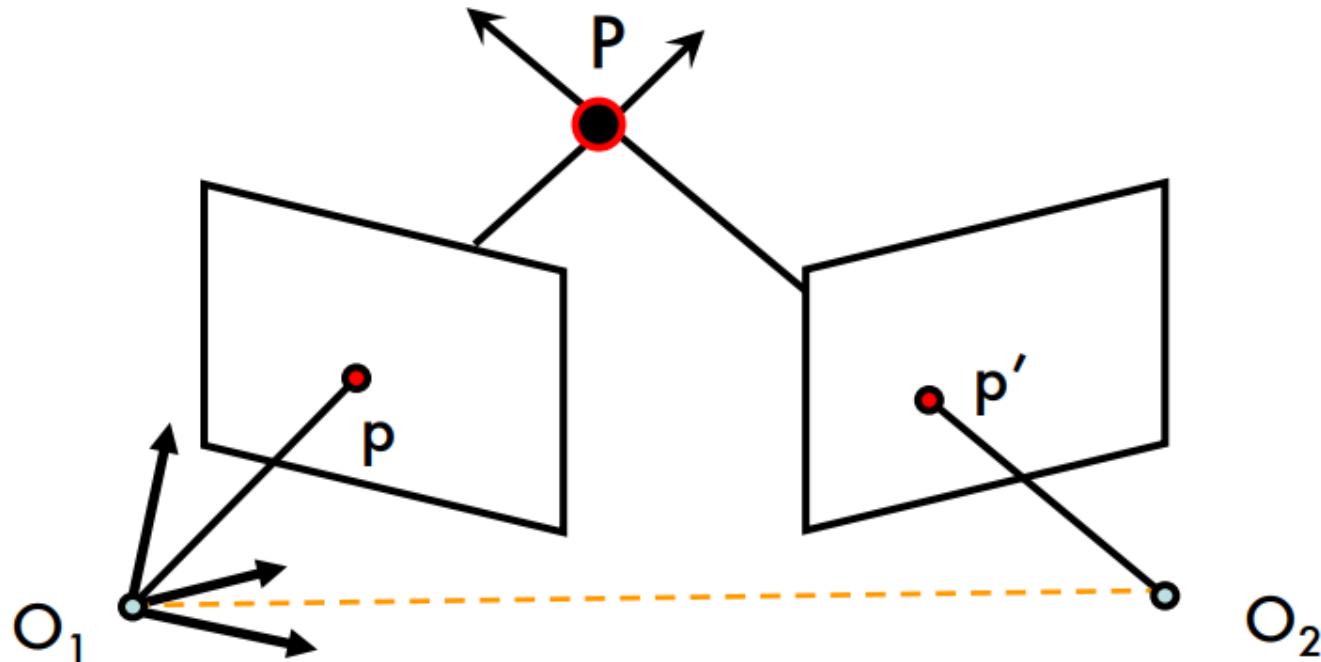
$$\mathbf{E} = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \mathbf{R}$$

$$p^T \mathbf{E} p' = 0$$

E = Essential Matrix

$$\mathbf{E} = [T_x] \cdot \mathbf{R}$$

Matrice fondamentale F



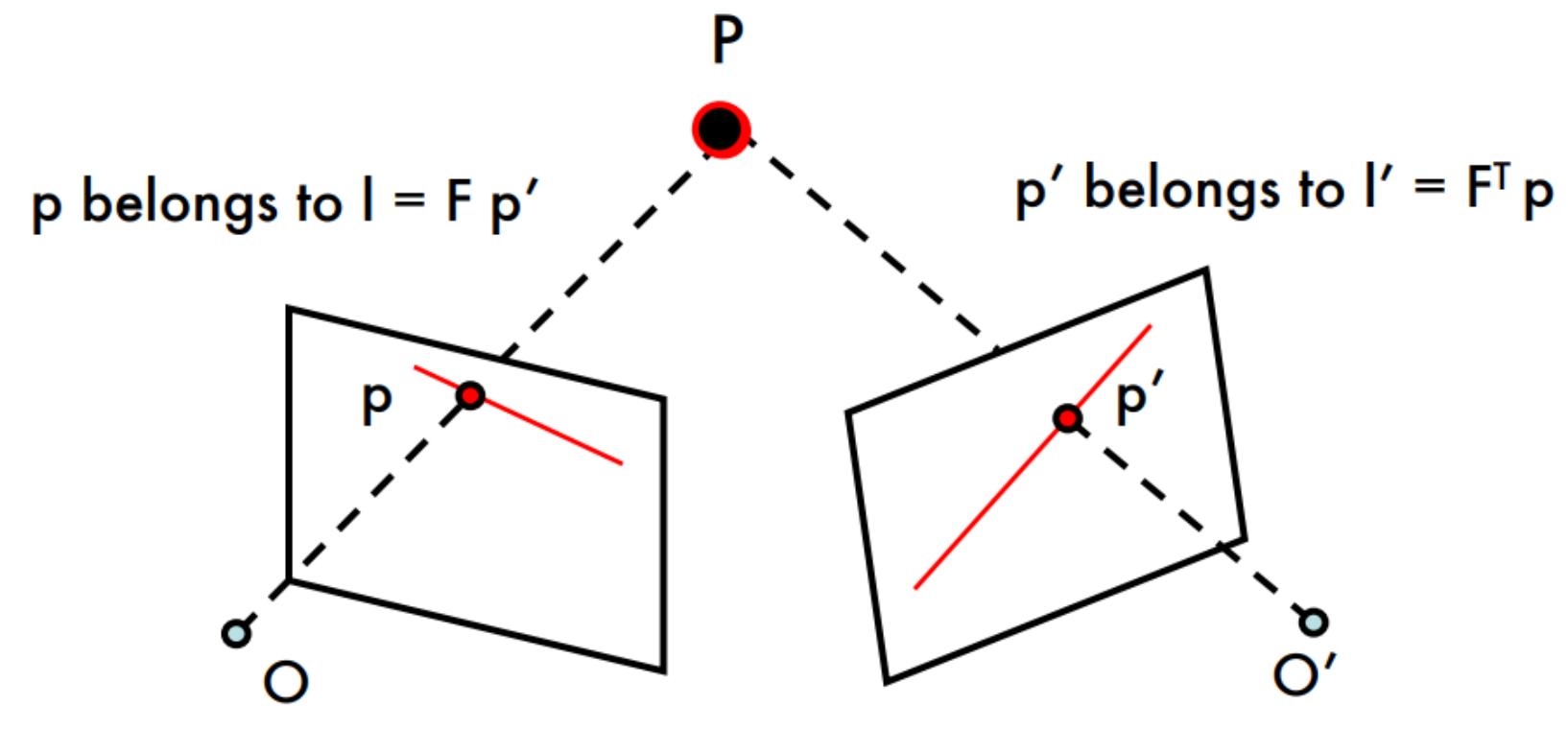
$$p^T F p' = 0$$

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

F = Fundamental Matrix

Appariement de points

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



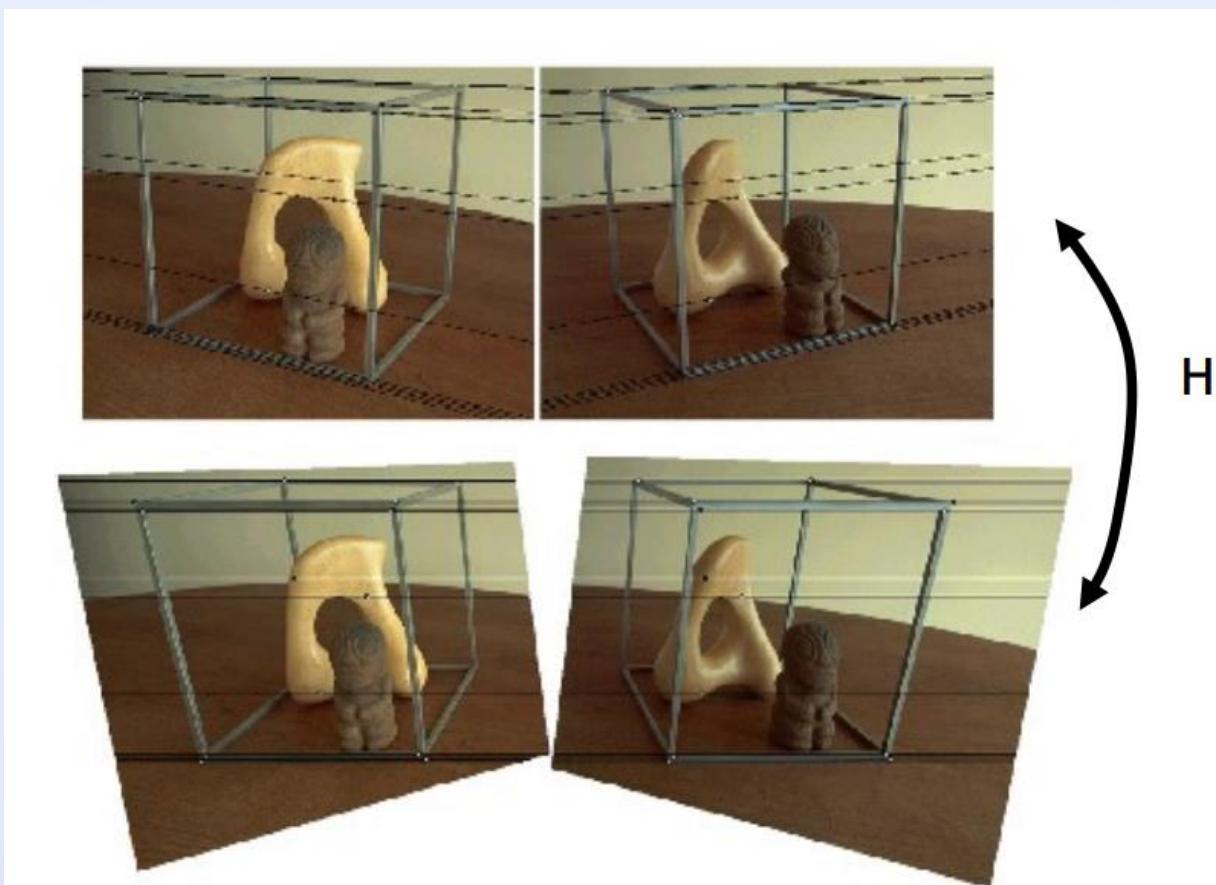


Rectification



Le fait de rendre deux images « parallèles »:

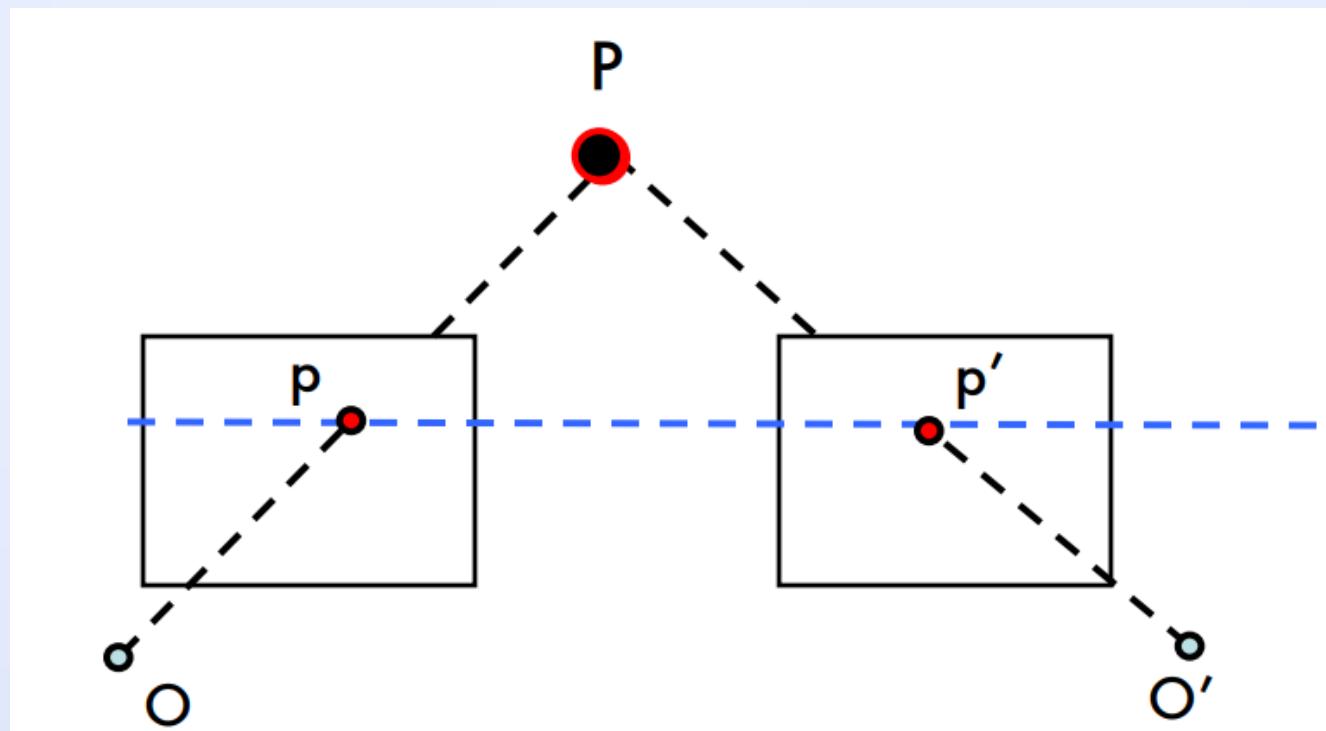
- Alignement des lignes épipolaires
 - Facilite l'appariement
 - Rends la triangulation plus facile



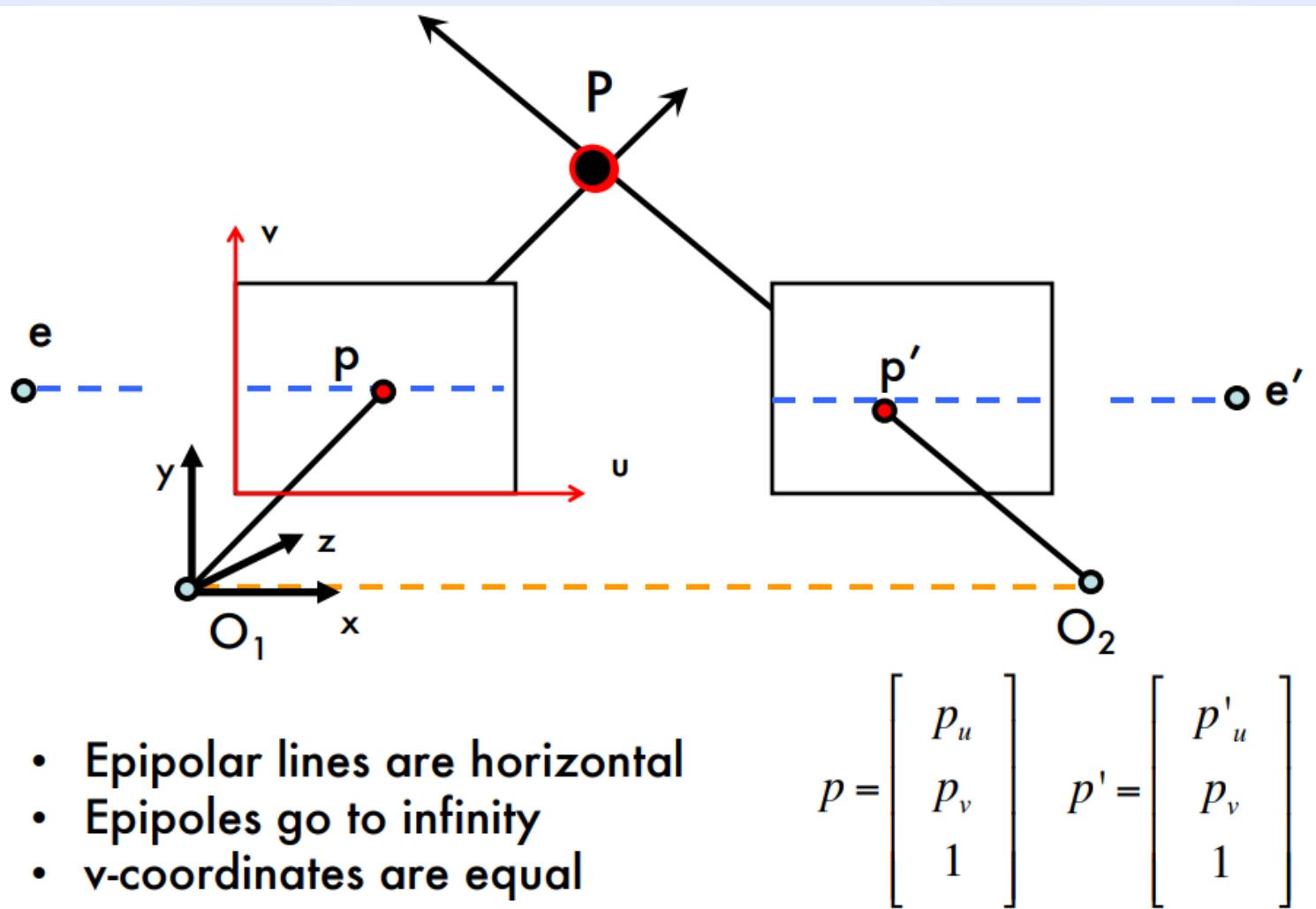


Appariement de points

- 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

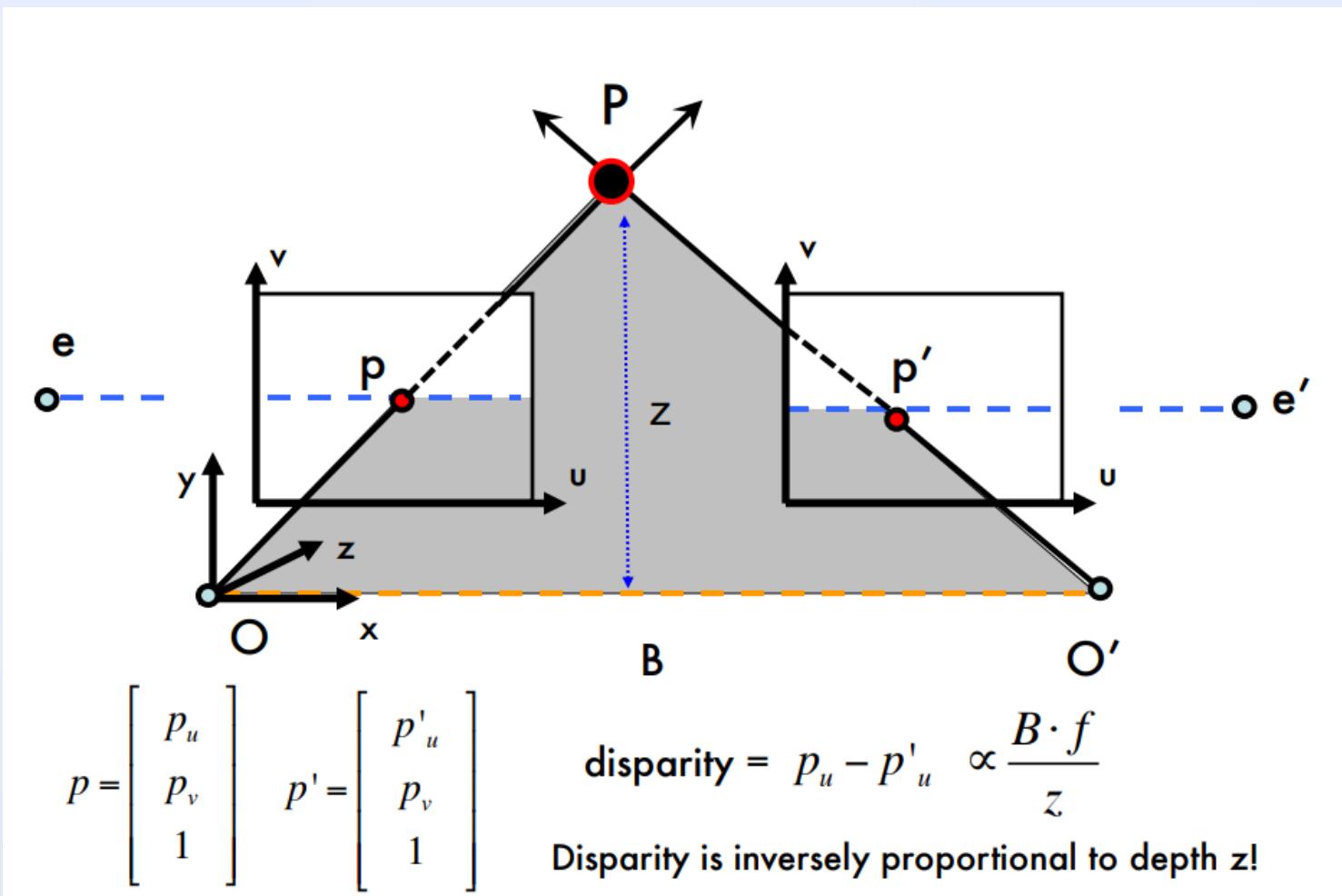


Plans images parallèles

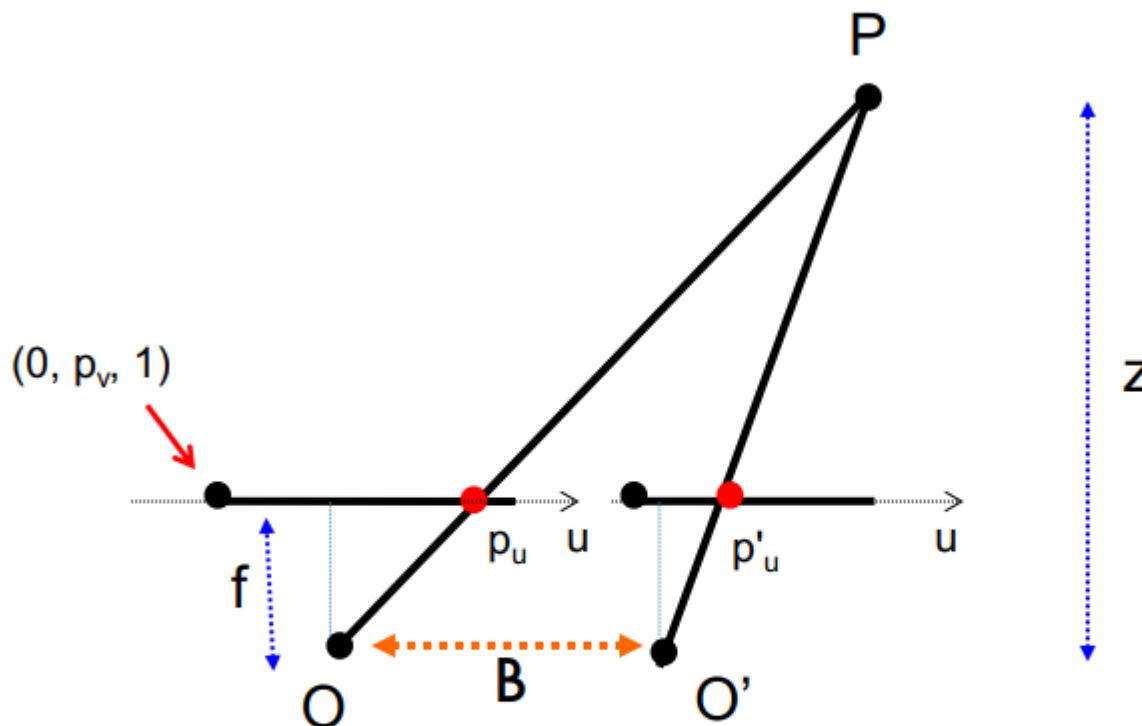


Disparité

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

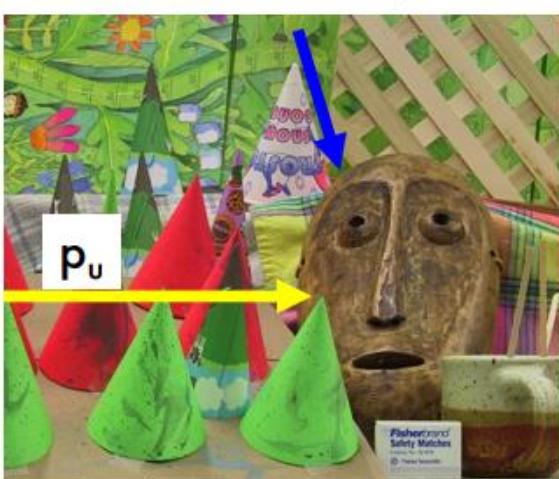


Calcul de la profondeur



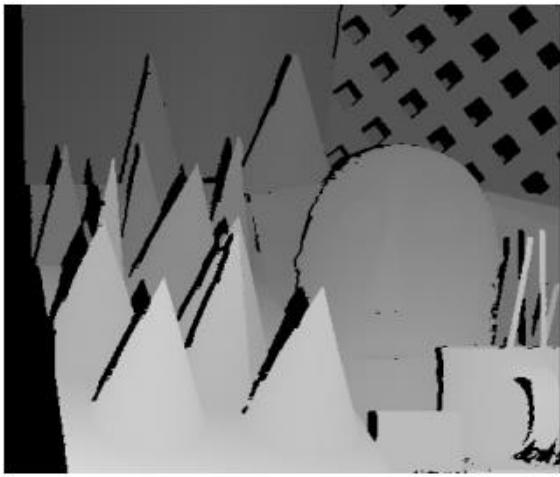
$$\text{disparity} = p_u - p'_u \propto \frac{B \cdot f}{z} \rightarrow z = \frac{B f_x}{d}$$

Image de disparité/carte de profondeur



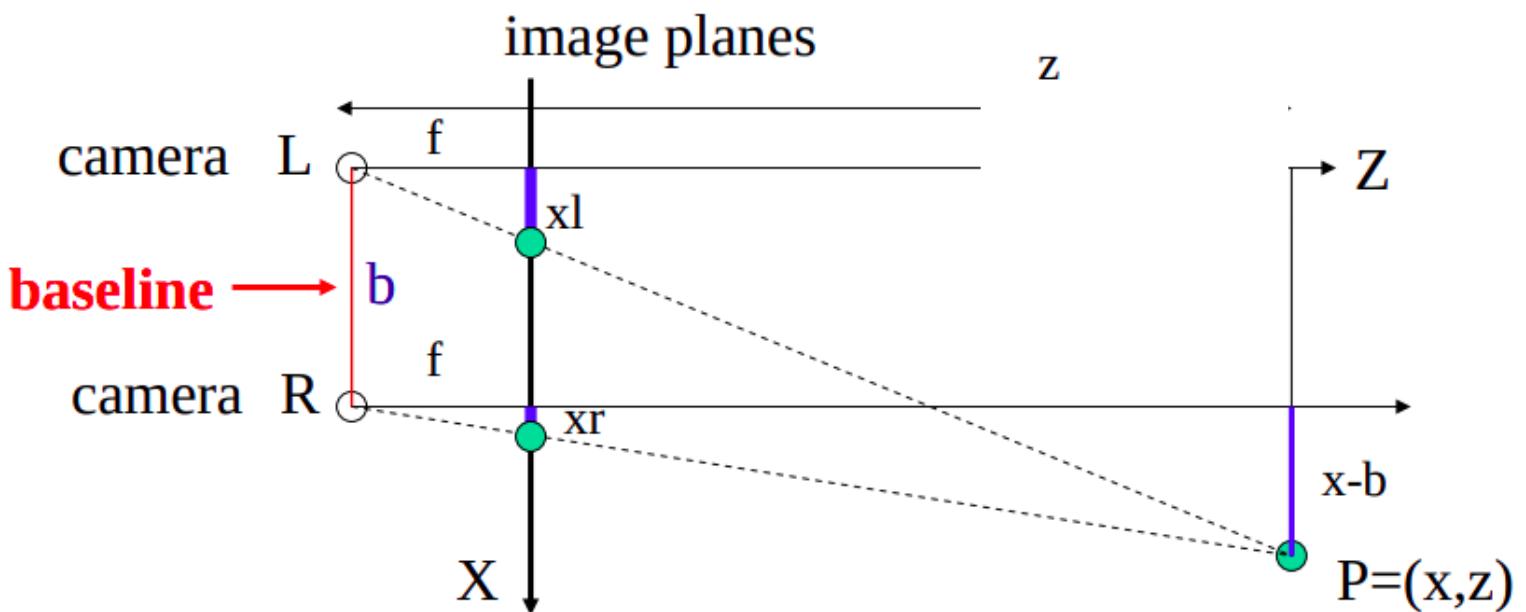
$$p_u - p'_u \propto \frac{B \cdot f}{z}$$

Stereo pair



Disparity map / depth map

Calcul des coordonnées 3D



- Théorème de Thalès:

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

$$\Rightarrow Z = \frac{B}{d} \cdot f_x$$

$$\Rightarrow X = x_L \cdot \frac{B}{d} = (x_R + d) \cdot \frac{B}{d}$$

$$\Rightarrow Y = y_L \cdot \frac{B}{d} \cdot \frac{f_x}{f_y} = y_R \cdot \frac{B}{d} \cdot \frac{f_x}{f_y}$$



Méthodes de corrélation

Comment faire pour trouver les points correspondants?

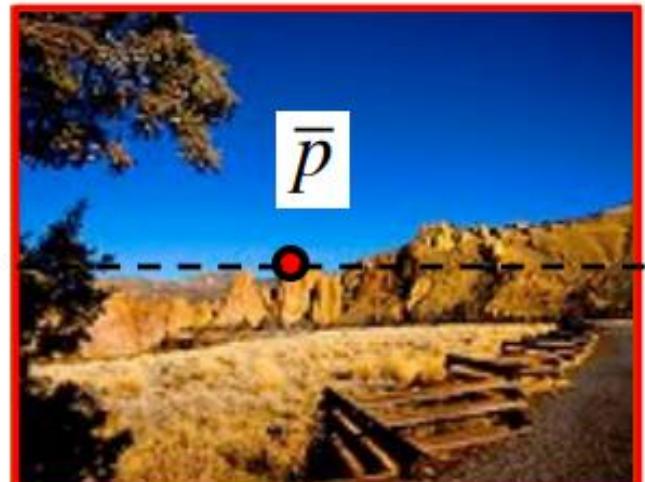


image 1

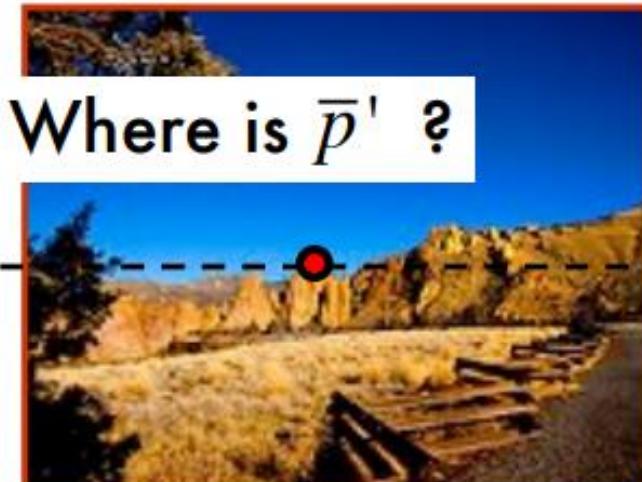
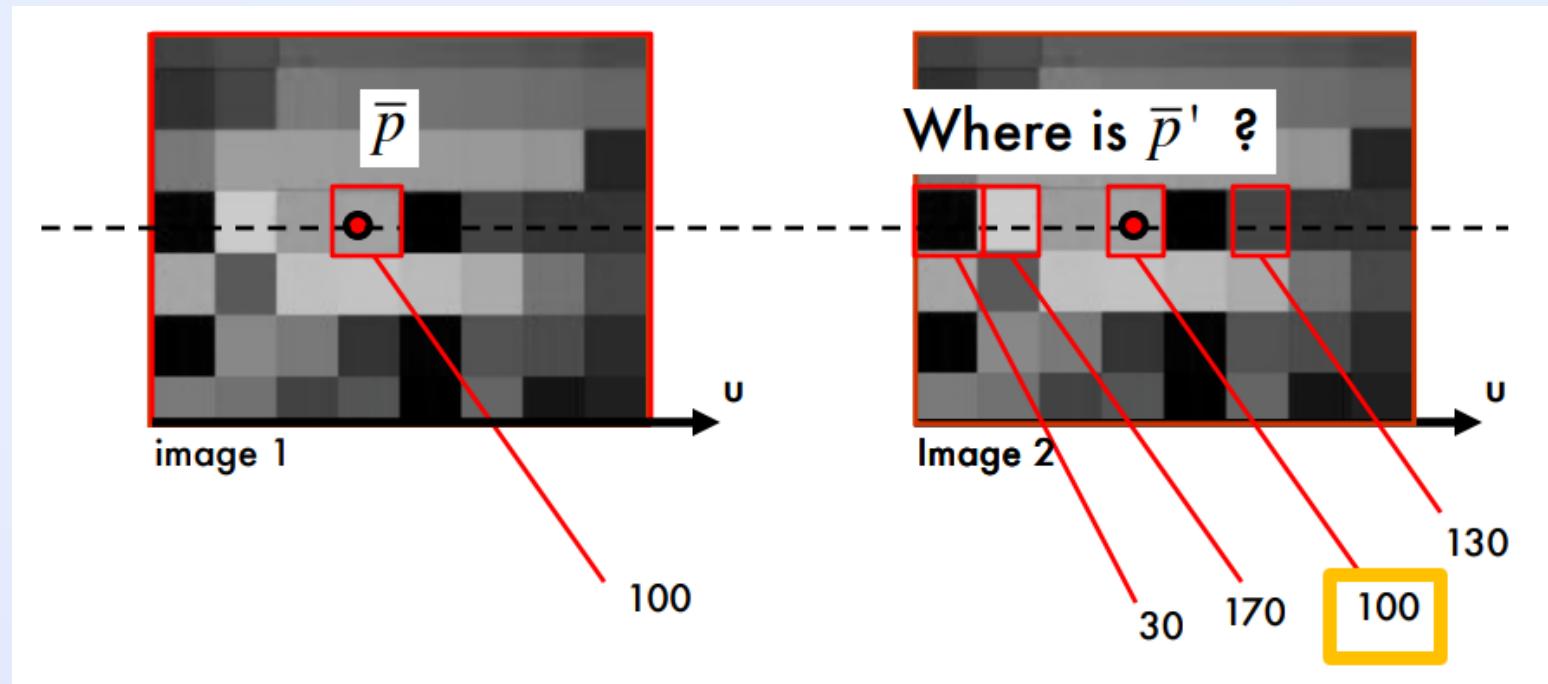


Image 2



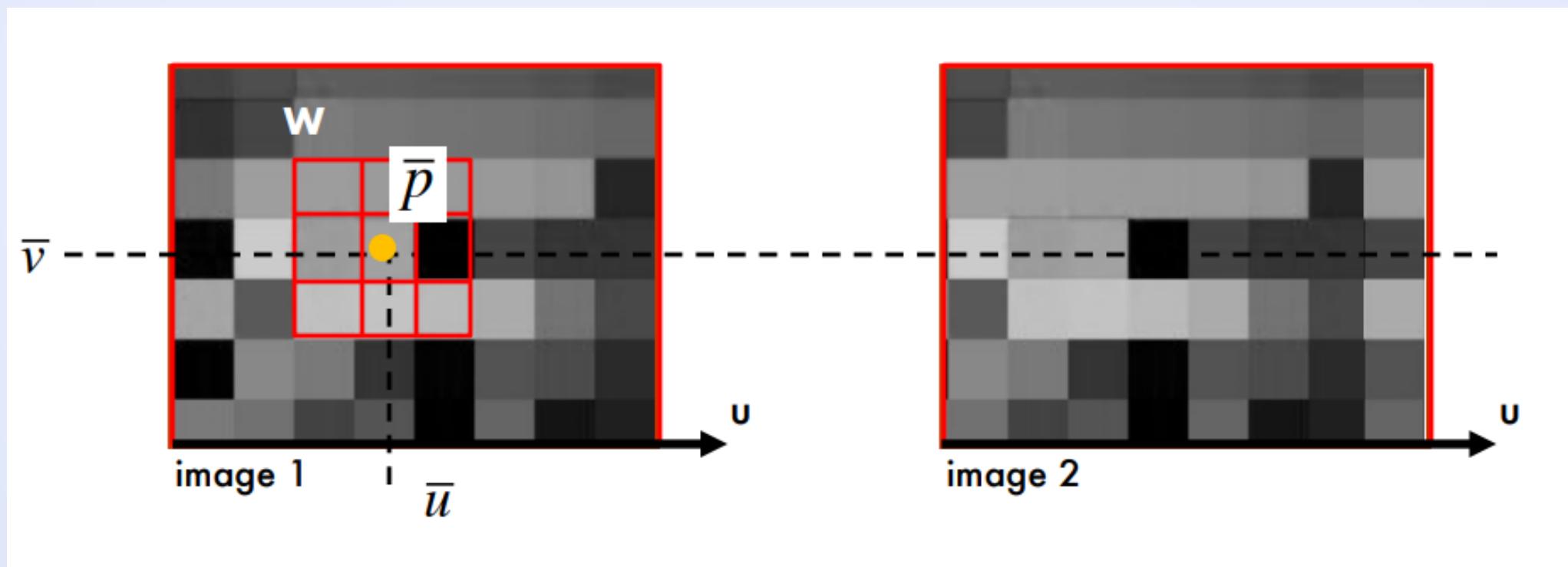
Corrélation par valeur

Recherche de valeur identique

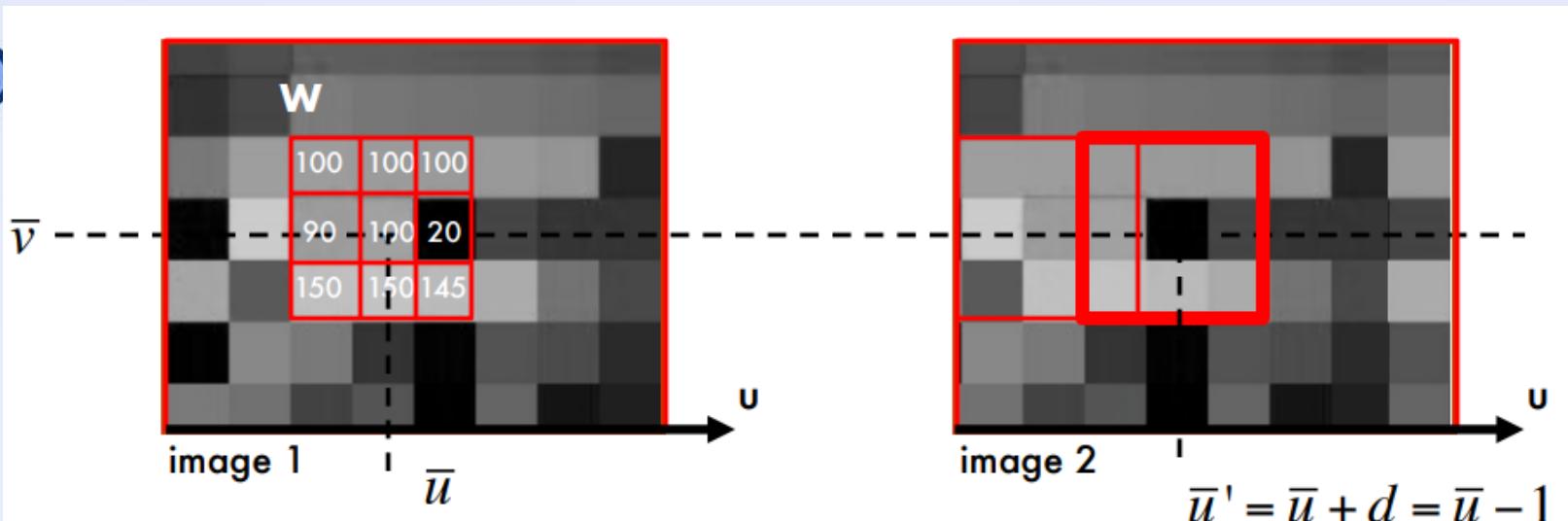


Fenêtre de corrélation

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



Fenêtre de corrélation



Example: **W** is a 3×3 window in red

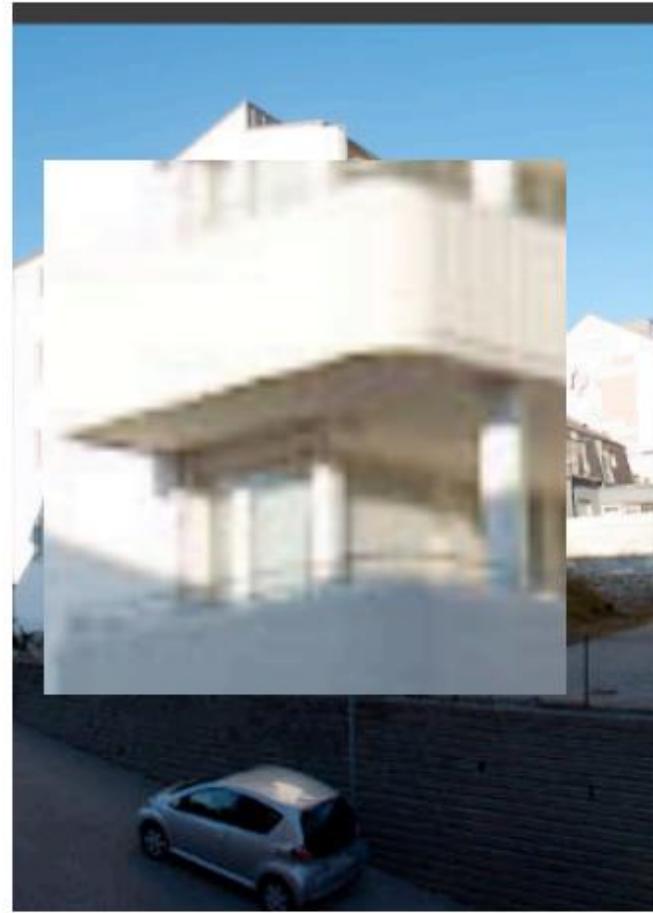
w is a 9×1 vector

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

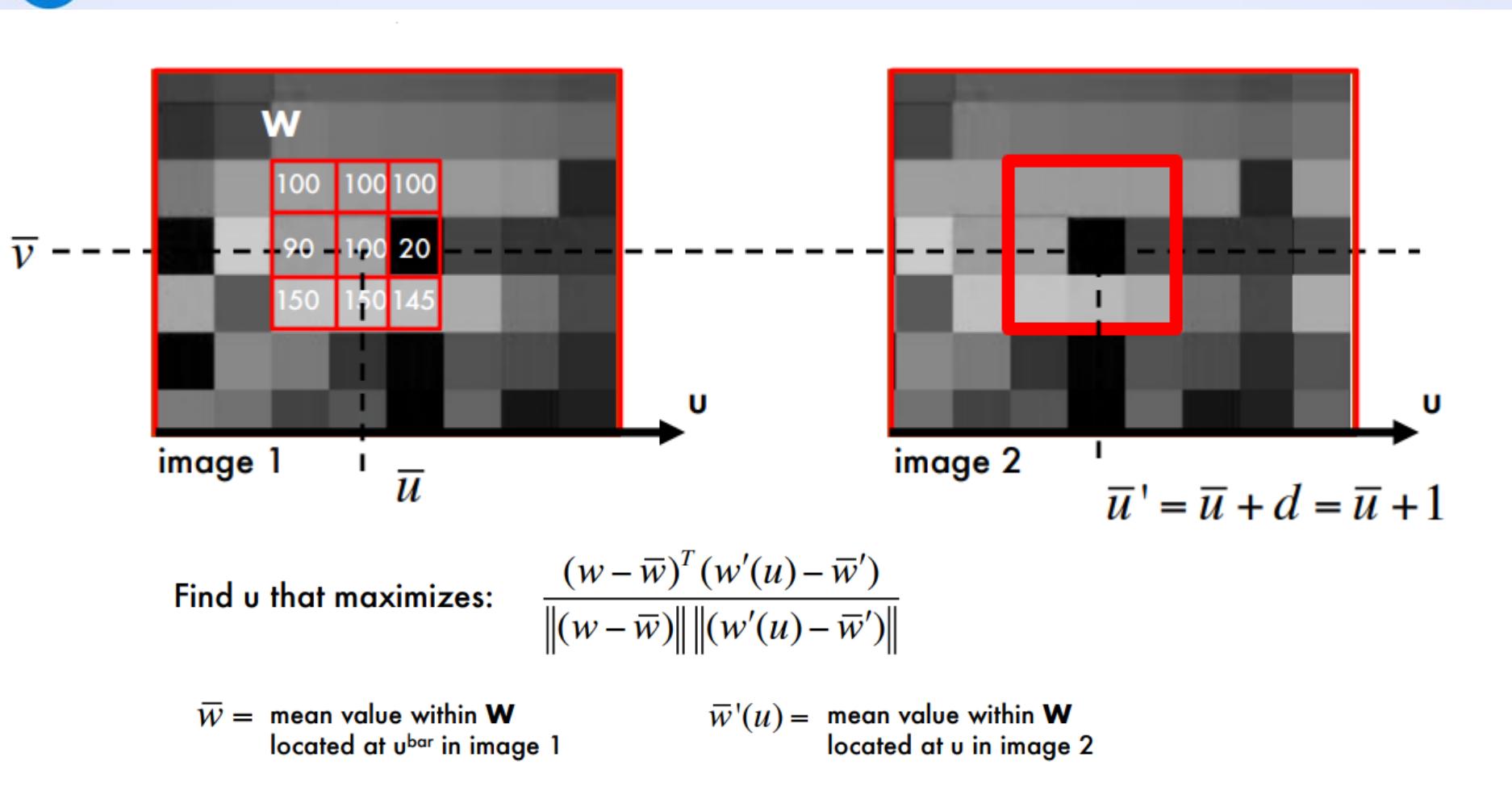
- Pick up a window **W** around $\bar{p} = (\bar{u}, \bar{v})$
- Build vector **w**
- Slide the window **W** along $v = \bar{V}$ in image 2 and compute **w'** (u) for each u
- Compute the dot product $\mathbf{w}^T \mathbf{w}'(u)$ for each u and retain the max value

Fenêtre de corrélation

Sensible aux différences d'exposition

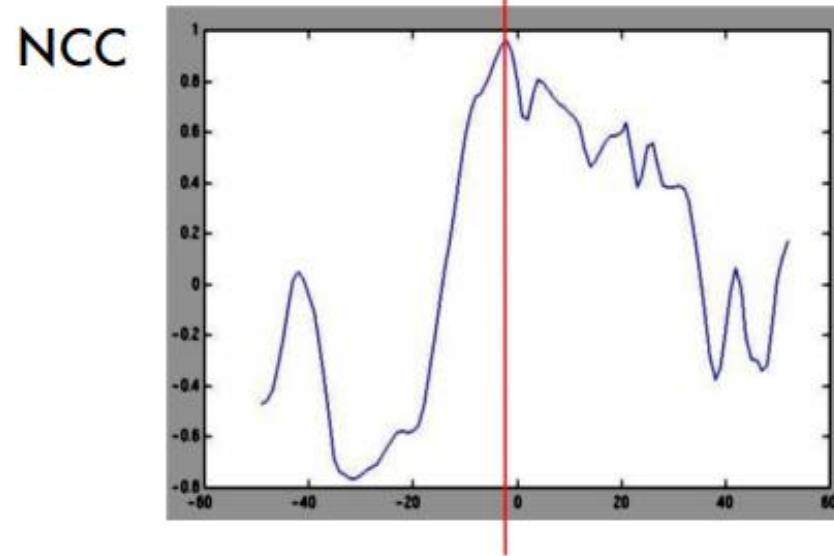


Normalized cross-correlation



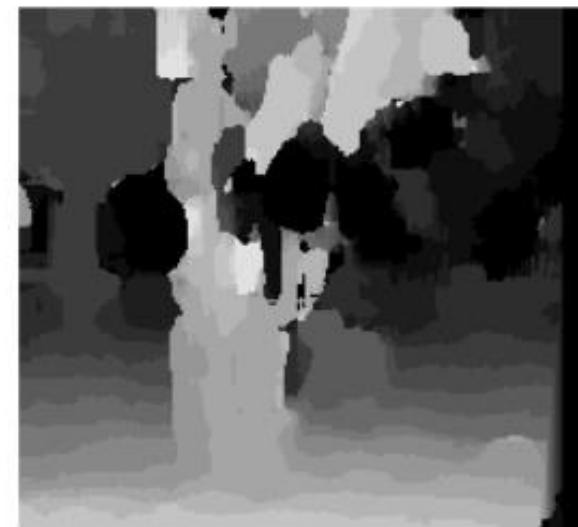


• Normalized cross-correlation



Taille de la fenêtre de corrélation

- Petite fenêtre:
 - + Plus de détail
 - Plus de bruit
- Grande fenêtre:
 - + Moins de bruit
 - Moins de détails

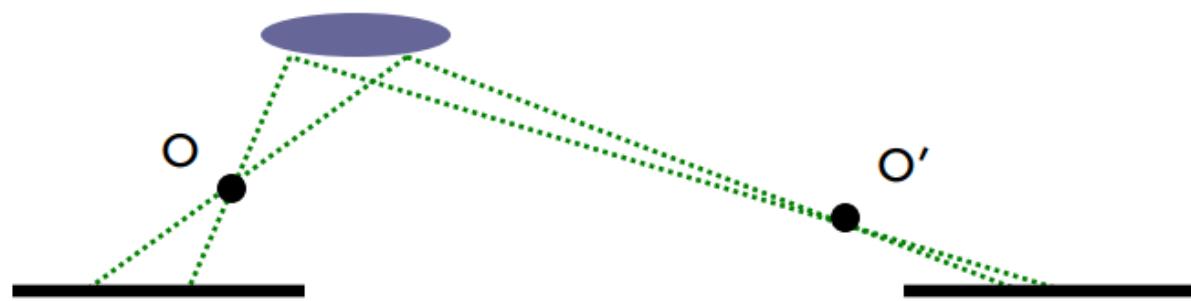


Window size = 3

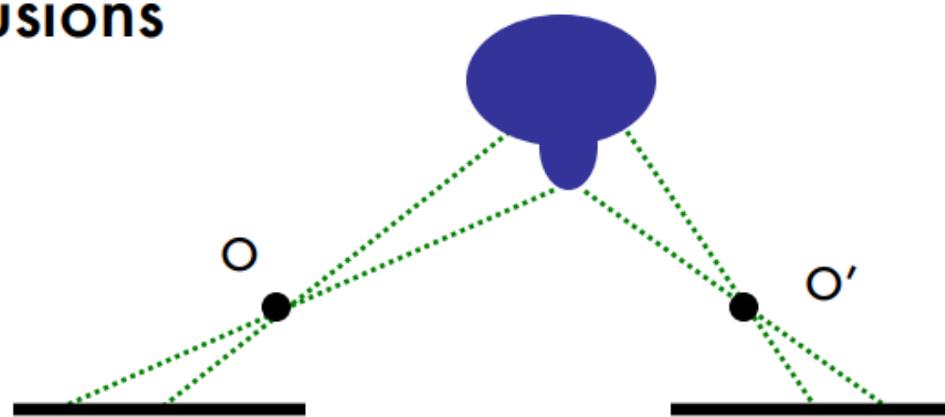
Window size = 20

Problèmes de la stéréo-vision

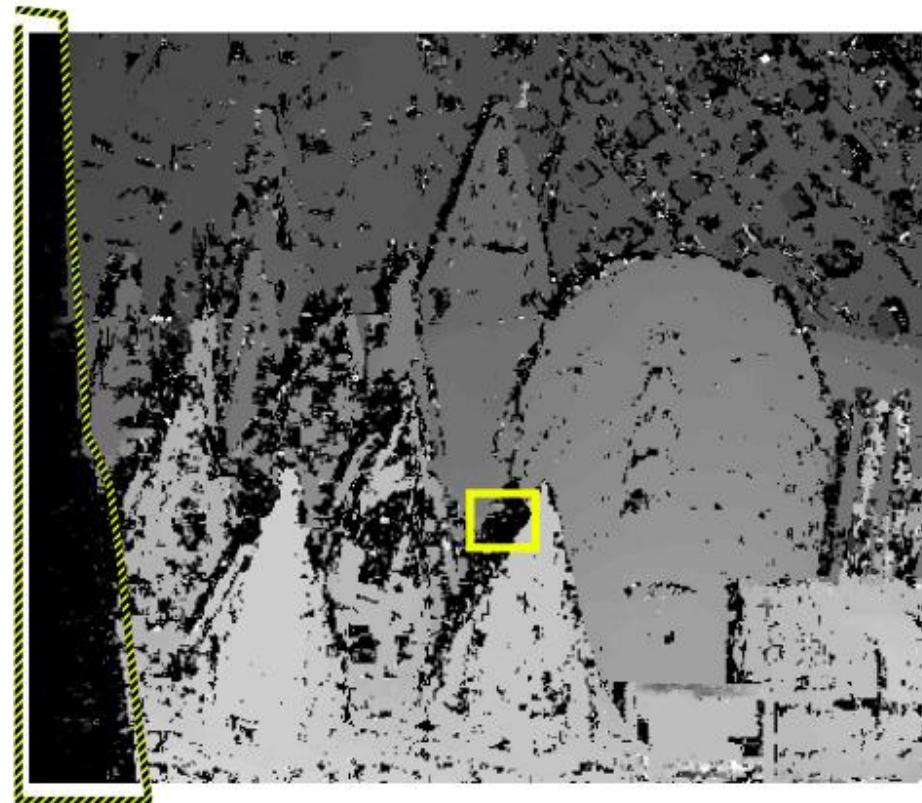
- Fore shortening effect



- Occlusions



Occlusions



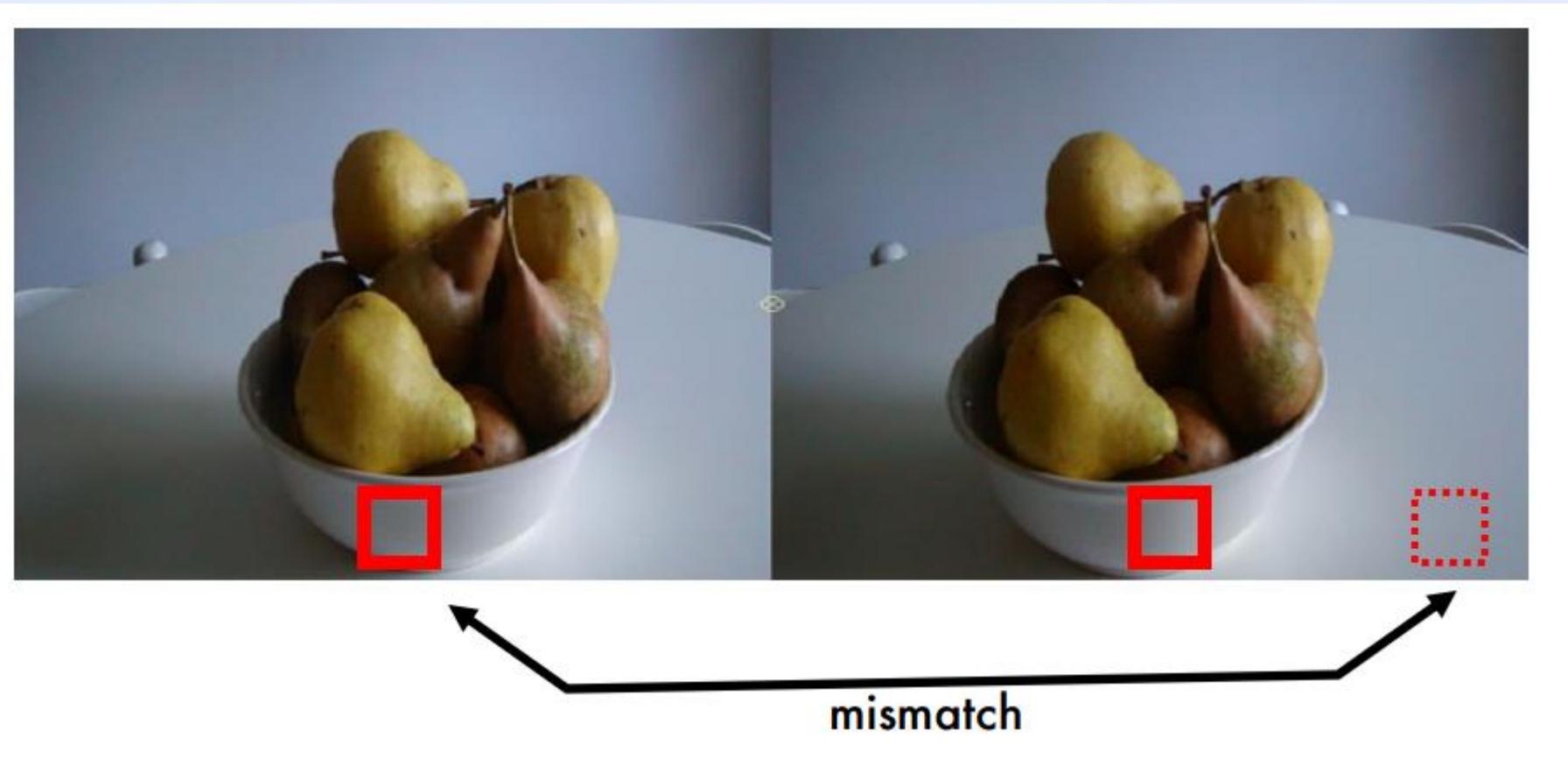
Left image



Right image

Problèmes de la stéréo-vision

Régions homogènes et/ou peu texturées



Problèmes de la stéréo-vision

Patterns répétitifs





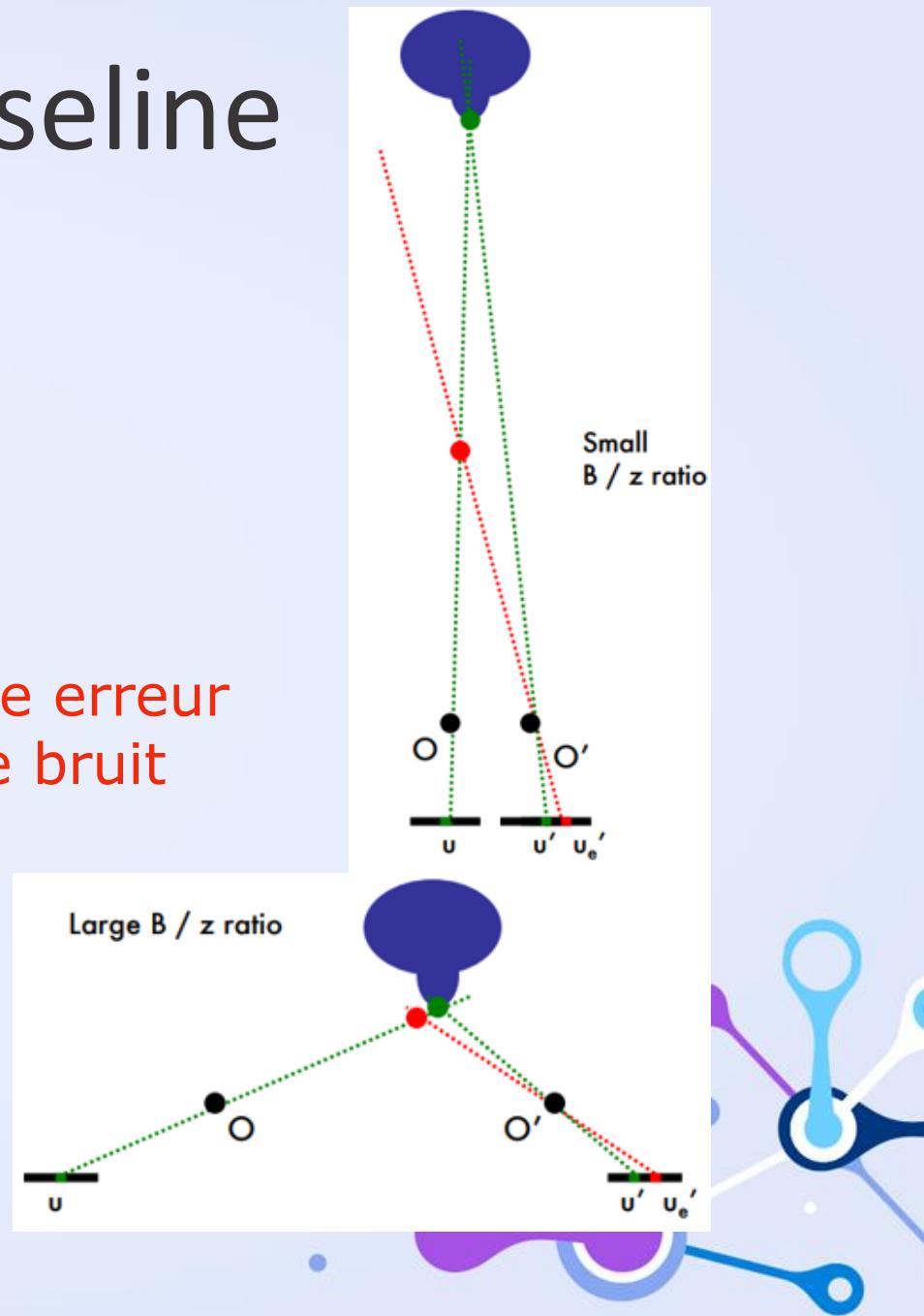
Influence de la Baseline

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



Disparité entière

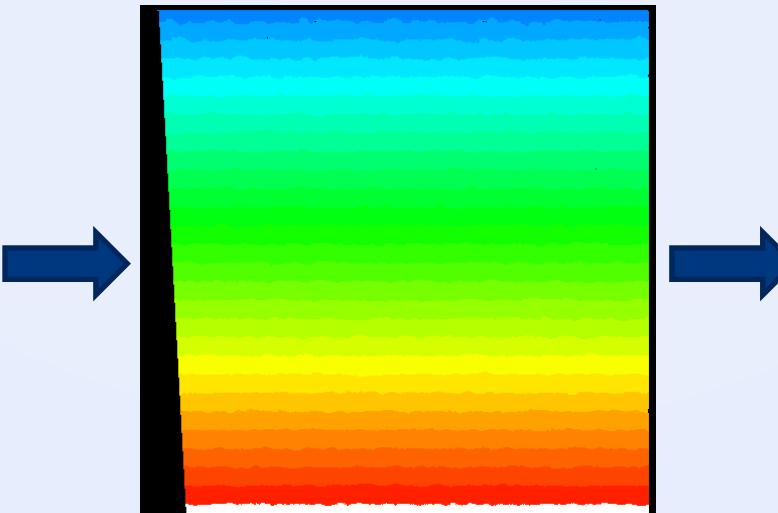
Le processus de stéréo-corrélation permet uniquement de calculer des valeurs entières de disparité.

=> Il est nécessaire de réaliser une interpolation pour lisser les disparités.

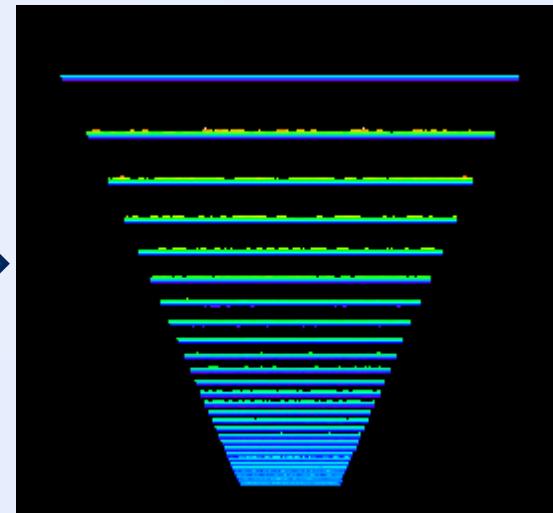
Exemples sans interpolation sous-pixellique:



• Image d'origine



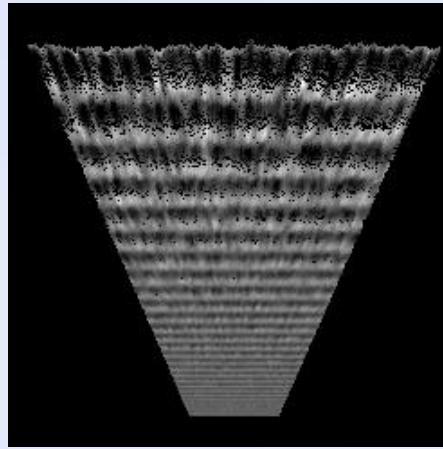
Carte de disparité



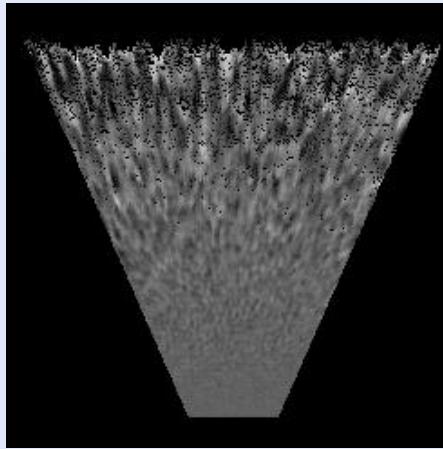
MNT

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.



Sym V



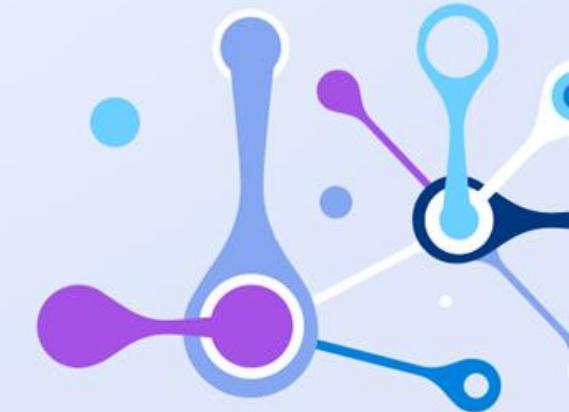
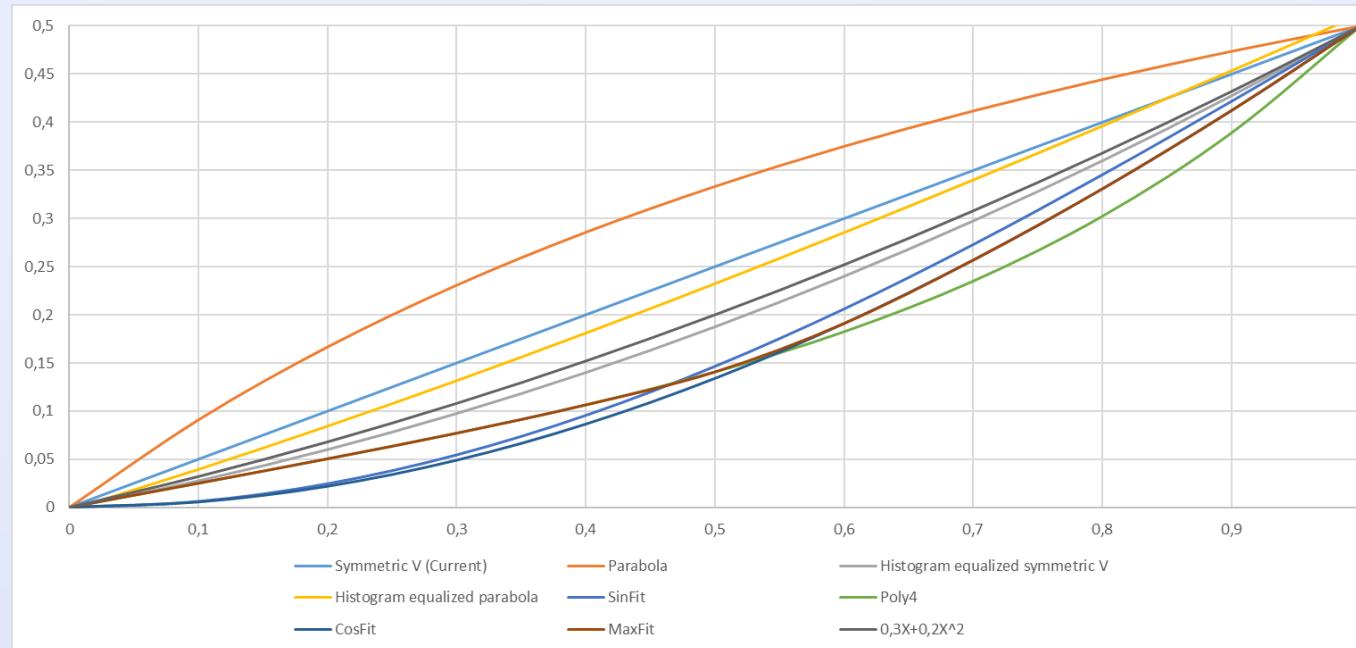
Hist. eq. Sym V

Interpolation sous-pixellique

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

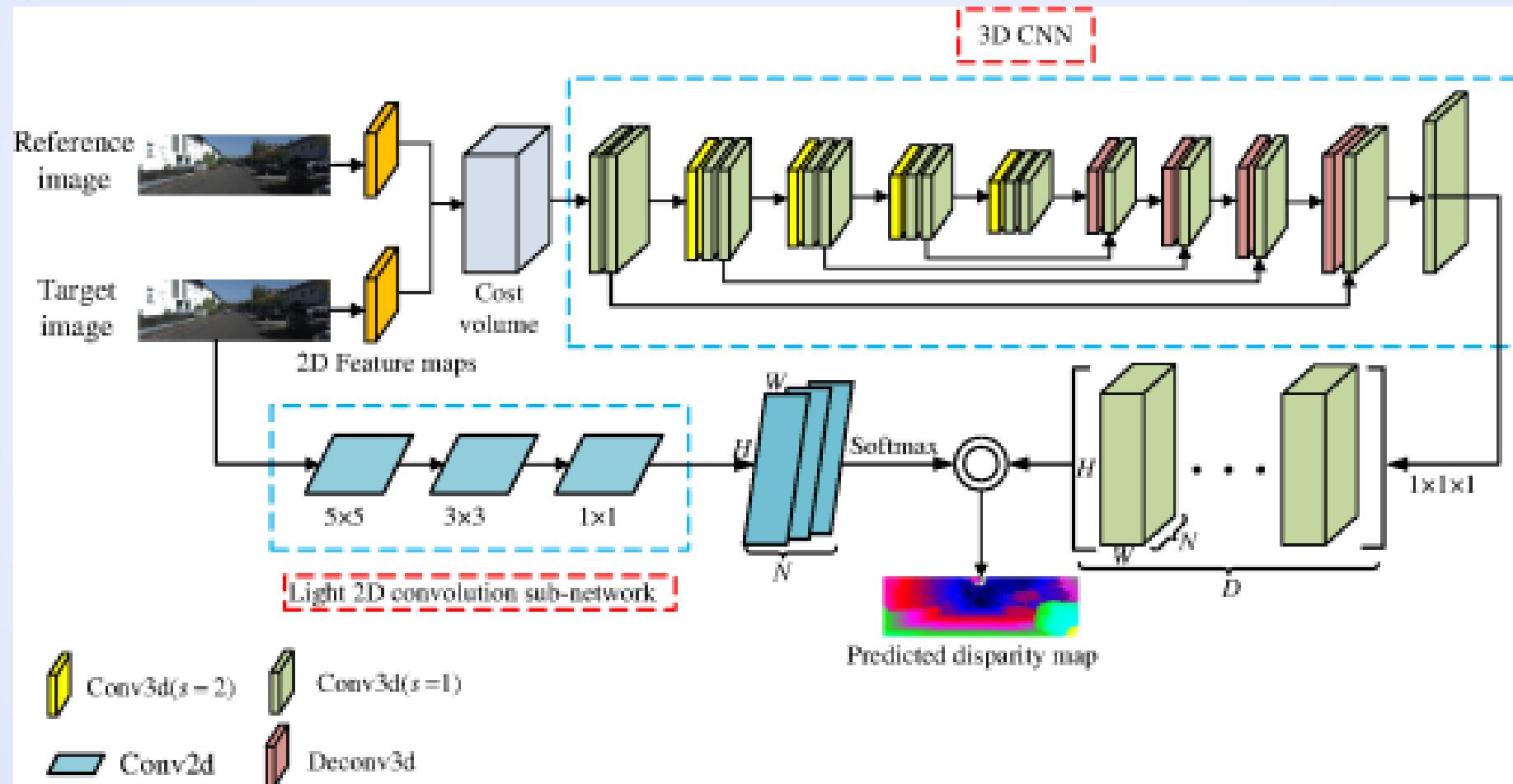
$$d_{Subpix} = \begin{cases} d_{int} - 0.5 + f\left(\frac{l_d}{r_d}\right), & \text{if } l_d \leq r_d \\ d_{int} + 0.5 - f\left(\frac{r_d}{l_d}\right), & \text{otherwise} \end{cases}$$

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



Deep-Learning

Réseaux de neurones convolutifs

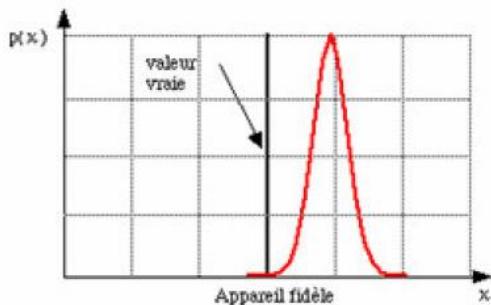


Stéréovision

Etat de l'art:

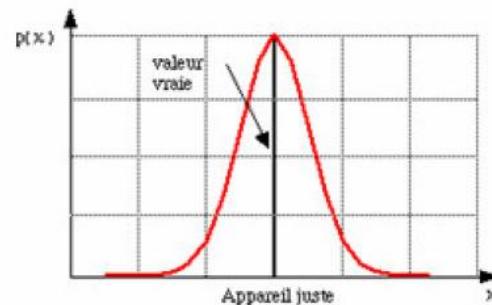
<https://vision.middlebury.edu/stereo/>

Evaluation:



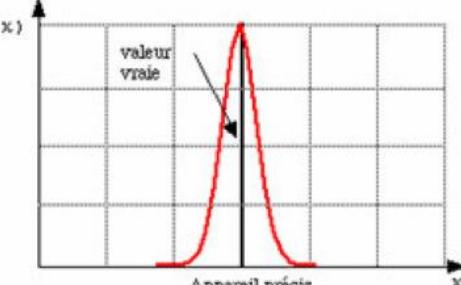
Fidélité : analysée par l'écart-type

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



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.