

# Vision par ordinateur - Reconnaissance Visuelle

## Cours 7: Un peu de géométrie

### Formation de l'image - Modèle de caméra - Stéréovision - Reconstruction 3D

Céline Hudelot - Mention IA - CentraleSupélec

2022-2023

# Ouvrages conseillés

- **Multiple View Geometry in Computer Vision**, R. Hartley, A. Zisserman
  - ▶ <http://www.robots.ox.ac.uk/~vgg/hzbook/>
  - ▶ Disponible sur le Drive
- Très bon cours de Princeton  
<http://3dvision.princeton.edu/courses/COS429/2015fa/>
- Cours Advanced Vision, MIT - 3 first courses  
<http://6.869.csail.mit.edu/sp21/schedule.html>
- Polycopié de P. Sturm  
<http://crowley-coutaz.fr/jlc/Courses/2008/GVR.V0/GVR-S6-sturm.pdf>

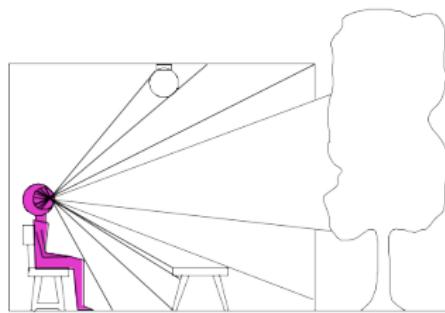
# Plan

- 1 Introduction
- 2 Projection perspective
- 3 Modèle de caméra
- 4 Retour sur la projection perspective
- 5 Calibration d'une caméra
- 6 Triangulation
- 7 Vision stéréoscopique
- 8 Structure from Motion
- 9 Conclusion

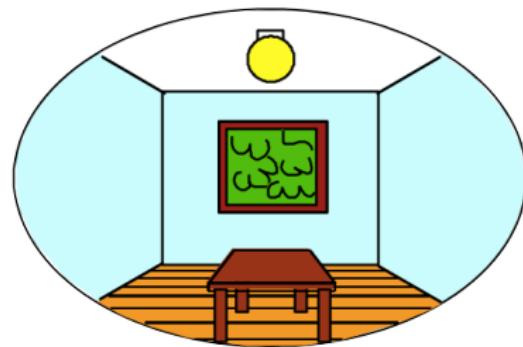
# Du monde réel à l'image

Formation d'une image

***3D world***



***2D image***



Point of observation

Source : S. E. Palmer

# Du monde réel à l'image

Passage d'une scène 3D (monde réel) à une scène 2D (image)



Distorsion d'images liée à la perte de profondeur et au point de vue

Question : comment retrouver la 3ème dimension ?

# Du monde réel à l'image

Une partie de l'information est perdue.

## Un rapide aperçu du problème

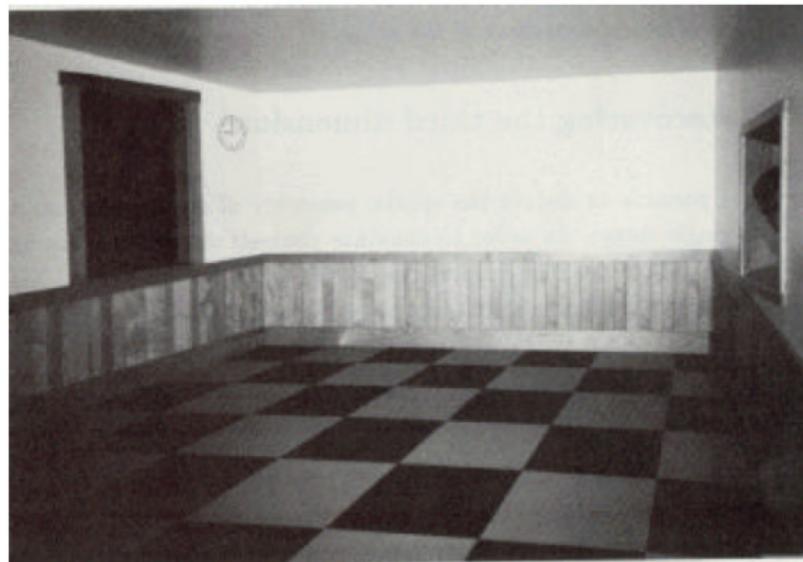


FIGURE – Source : E. Marchand, Irisa

# Du monde réel à l'image

Une partie de l'information est perdue.

## Un rapide aperçu du problème



source : E. Marchand, Irisa

FIGURE – Source : E. Marchand, Irisa

# Du monde réel à l'image

Une partie de l'information est perdue.

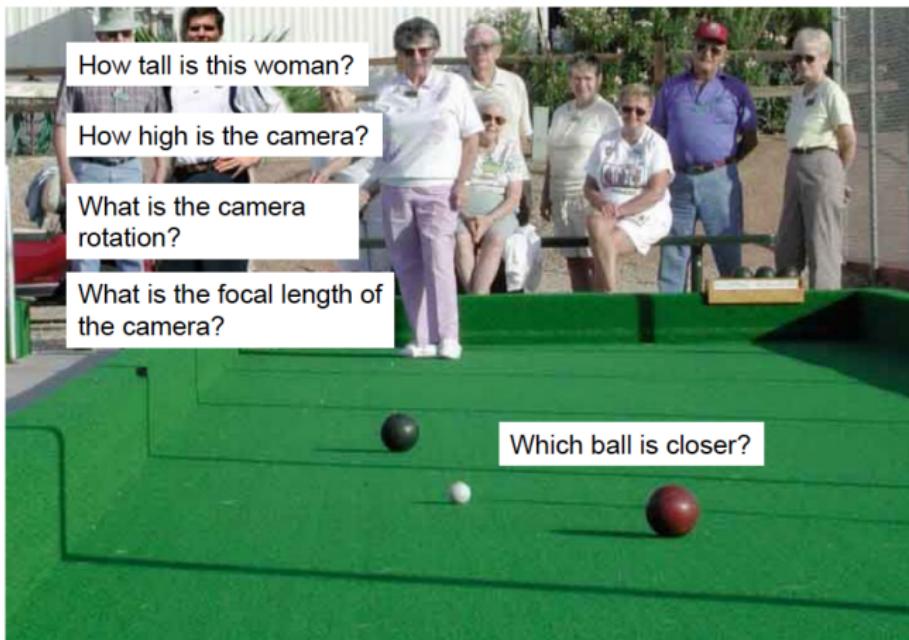
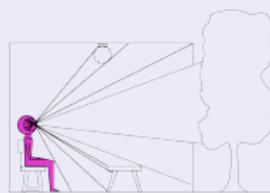


FIGURE – Source : J. Hays

# Du monde réel à l'image

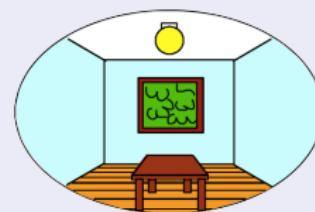
## Formation d'une image

*3D world*



Point of observation

*2D image*



Source : S. E. Palmer

Les images sont des projections du monde 3D dans un plan 2D :

- Projections : équations, géométrie.
- Modèles de caméras.

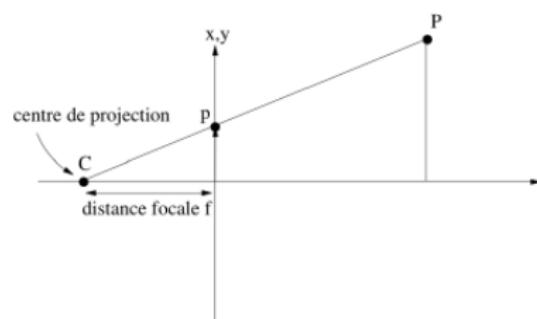
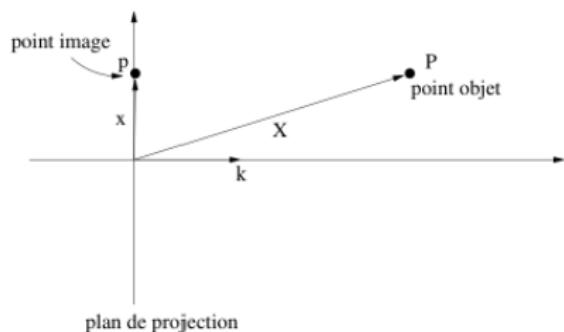
# Plan

- 1 Introduction
- 2 Projection perspective
- 3 Modèle de caméra
- 4 Retour sur la projection perspective
- 5 Calibration d'une caméra
- 6 Triangulation
- 7 Vision stéréoscopique
- 8 Structure from Motion
- 9 Conclusion

# Notion de projection

Plusieurs familles de projections :

- Projections parallèles dont la projection orthographique.
- Projections perspectives (effet visuel semblable à celui perçu par l'œil humain.)



**FIGURE – Source : E. Boyer, Irisa**

## Projection parallèle

La projection se fait dans le plan de projection suivant une direction.

- Projection orthographique : projection perpendiculaire au plan de projection.

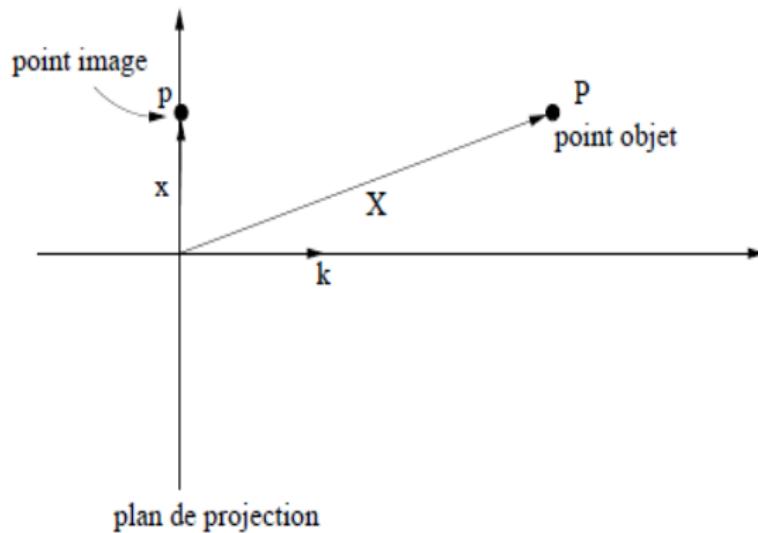
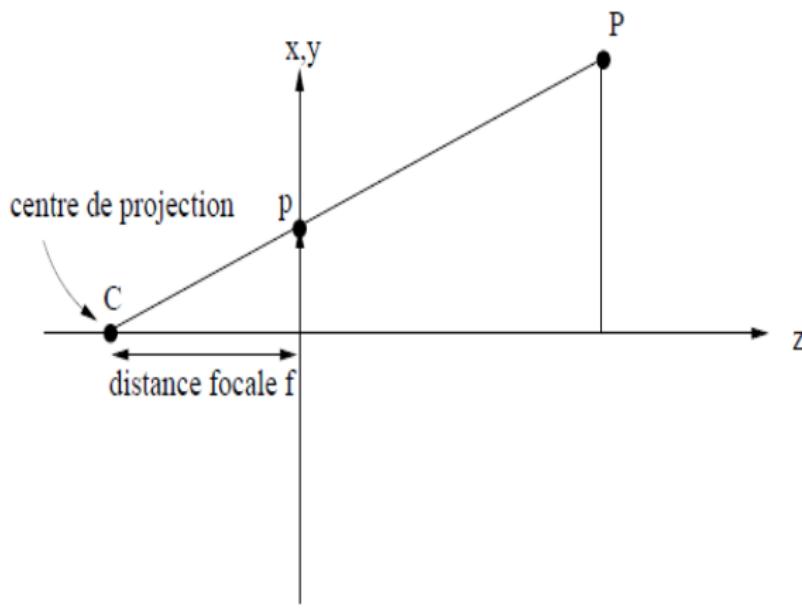


FIGURE – Source : E. Boyer, Irisa

# Projection perspective

La projection se fait suivant un centre de projection. L'origine du repère est placée au centre de projection et on l'oriente de sorte que le plan de projection ait pour équation  $z = f$  où  $f$  est la distance focale.



# Plan

- 1 Introduction
- 2 Projection perspective
- 3 Modèle de caméra
- 4 Retour sur la projection perspective
- 5 Calibration d'une caméra
- 6 Triangulation
- 7 Vision stéréoscopique
- 8 Structure from Motion
- 9 Conclusion

# Modèle de caméra

## Pourquoi ?

- Pouvoir effectuer des calculs numériques ou des raisonnements géométriques à partir d'images.
- ⇒ Besoin d'un modèle qui décrit comment le monde 3D se projette sur une image 2D, à partir d'une caméra.

## Modèle du sténopé (pin-hole)

- Modèle utilisé en vision par ordinateur.
- Un bon compromis entre la simplicité des équations associées et la proximité aux phénomènes physiques modélisés.
- Modèle sténopé (pinhole, trou d'épingle) : représente une projection perspective.

# Modèle de caméra : intuition

## Formation de l'image

Elle dépend :

- de l'éclairage de l'environnement<sup>a</sup>,
- des propriétés photométriques des surfaces,
- des propriétés géométriques des surfaces
- de l'optique de la caméra et des propriétés du capteur.

a. Voir les cours en informatique graphique comme par exemple

<https://cs.wellesley.edu/~cs307/readings-s21/Material-Lighting.html>

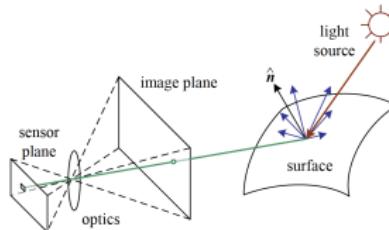


FIGURE – Source : Szeliski book

# Modèle de caméra : intuition

A propos de la lumière ambiante !

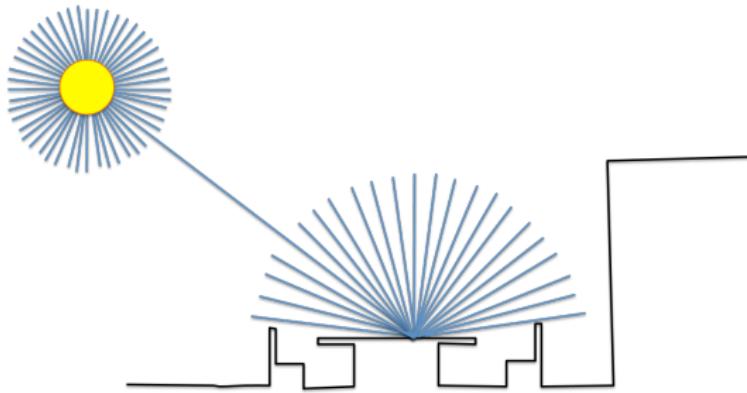


FIGURE – Source : MIT course

# Modèle de caméra : intuition

A propos de la lumière ambiante !

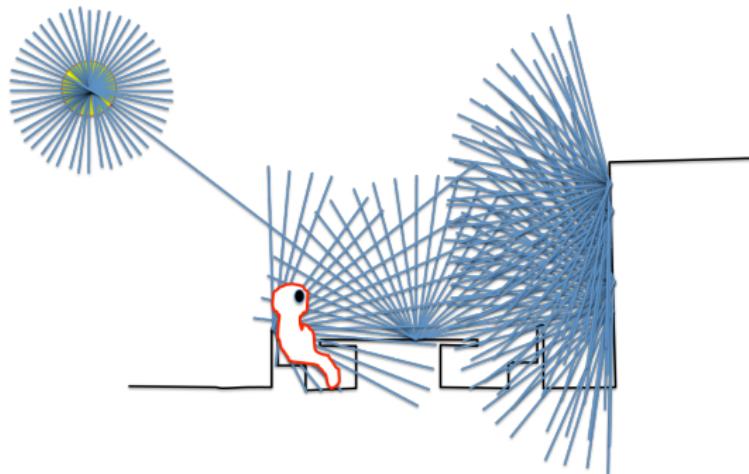


FIGURE – Source : MIT course

# Modèle de caméra : intuition

A propos de la lumière ambiante ! Peut-on se contenter de mettre une feuille blanche en face de l'objet pour générer une image ?

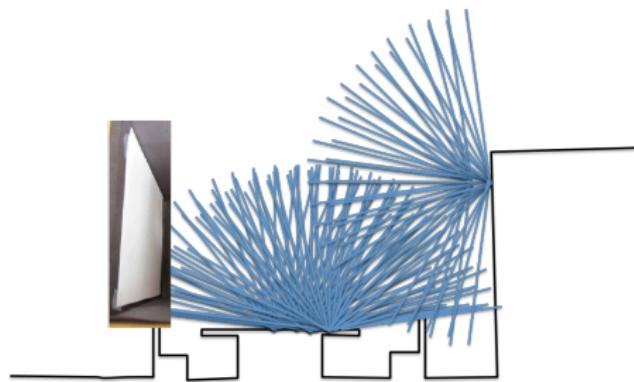


FIGURE – Source : MIT course

# Modèle de caméra : intuition

## Conception d'un appareil photo

- Mettre un film devant l'objet
- Va-t-on obtenir une image satisfaisante ?

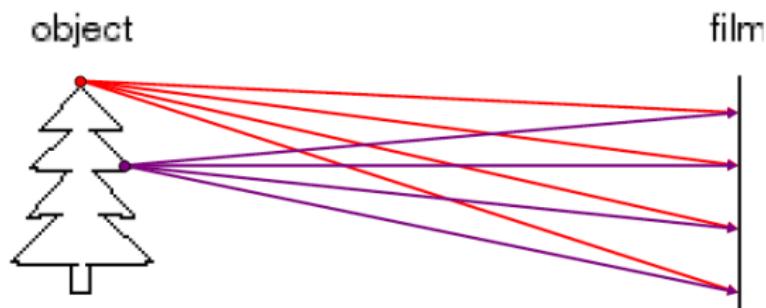


FIGURE – Source : S. Seitz

# Modèle de caméra : intuition

Va-t-on obtenir une image satisfaisante ? L'expérience peut-être faite facilement.

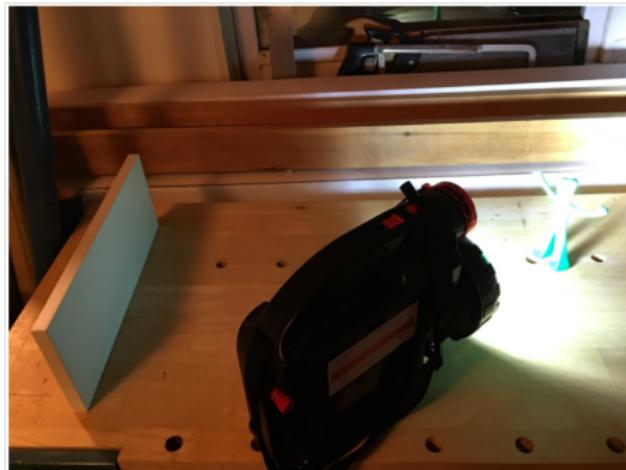


FIGURE – Source : MIT course

# Modèle de caméra : intuition

La réponse est **NON !**



FIGURE – Source : MIT course

# Modèle de caméra : intuition

Conception d'un appareil photo : idée simple

Les rayons lumineux provenant de différentes parties de la scène arrivent au même point sur le film.

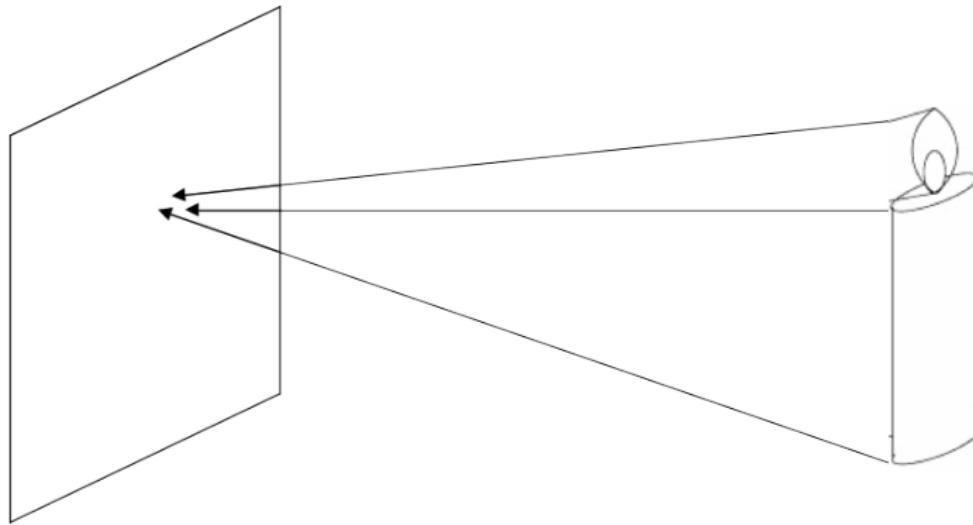


FIGURE – Source : Forsyth and Ponce

# Modèle de caméra : intuition

## Conception d'un appareil photo

On ajoute un barrage pour bloquer la plupart des rayons :

- Reduction du flou.
- Le trou est l'ouverture (*aperture*).

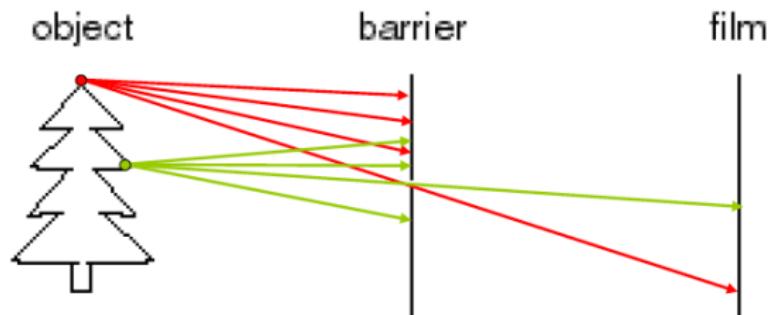


FIGURE – Source : S. Seitz

# Modèle de caméra : intuition

Retour à la petite expérience du MIT !

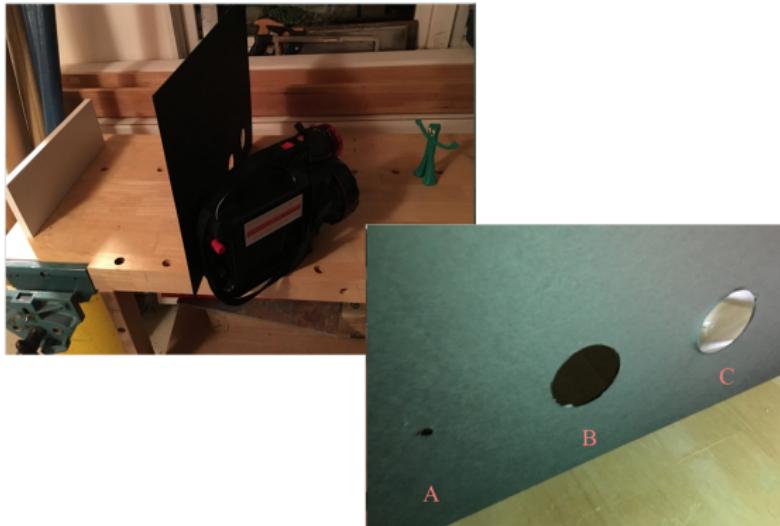


FIGURE – Source : MIT course

# Modèle de caméra : intuition

Retour à la petite expérience du MIT !



FIGURE – Source : MIT course

# Modèle de caméra : intuition

## Conception d'un appareil photo

Sténopé ou pin-hole camera : ne permet aux rayons lumineux provenant d'un point dans la scène que d'atteindre un seul point dans l'image.

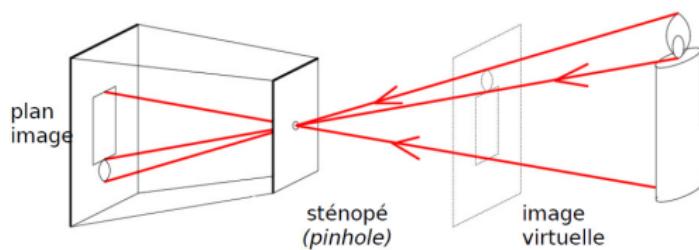


FIGURE – Source : A. Boucher

# Modèle de caméra : intuition

The camera obscura  
The pinhole camera

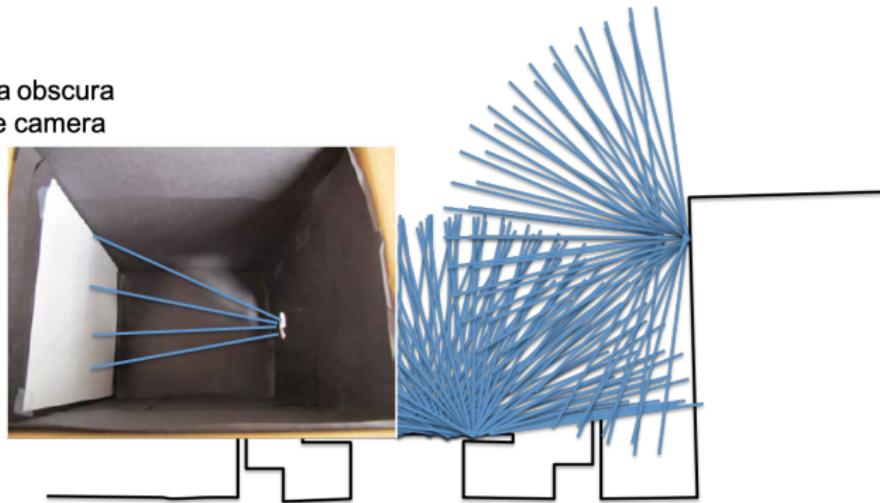


FIGURE – Source : MIT course

# Modèle de caméra : un peu d'histoire

## Camera Obscura

Leonard de Vinci (1452-1519) : premiers écrits sur la chambre obscure.

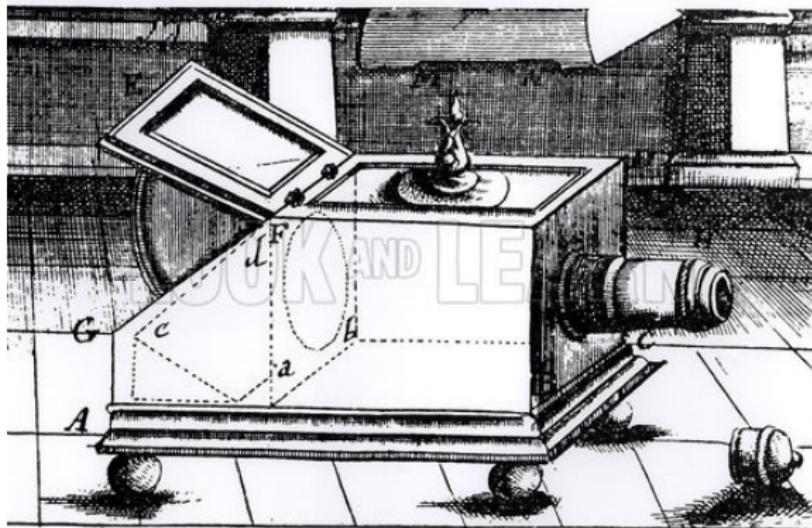


# Modèle de caméra : un peu d'histoire

Premier appareil photo portable

Johann Zahn (1685)

## THE HISTORY OF THE CAMERA OBSCURA



5. Johann Zahn. Reflex box camera obscura, 1685

# Modèle de caméra : un peu d'histoire

Première photographie

Joseph Nicéphore Niépce (1826)



# Modèle de caméra : Sténopé



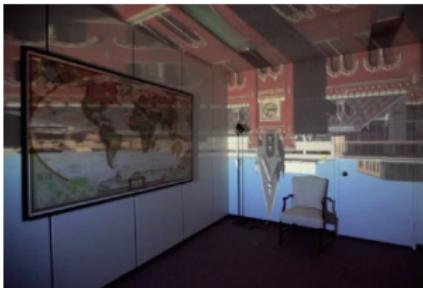
# Modèle de caméra : Sténopé



# Modèle de caméra : Sténopé

Quelques expériences intéressantes autour de ce modèle :

- <https://www.youtube.com/watch?v=gvzpu0Q9RTU>
- <http://people.csail.mit.edu/torralba/research/accidentalcameras/>



a) Input (occluder present)      b) Reference (occluder absent)

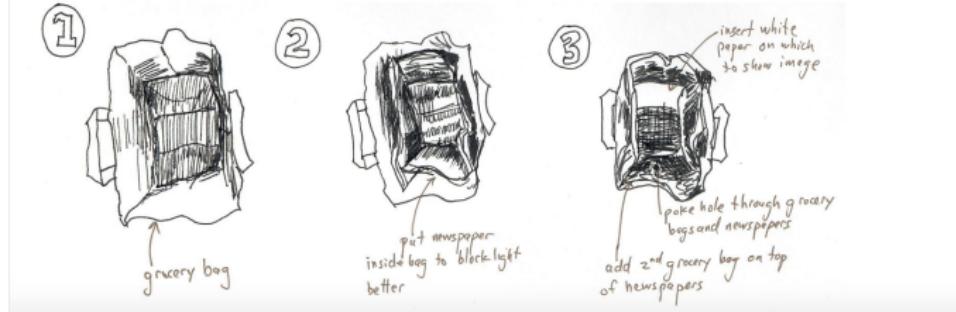
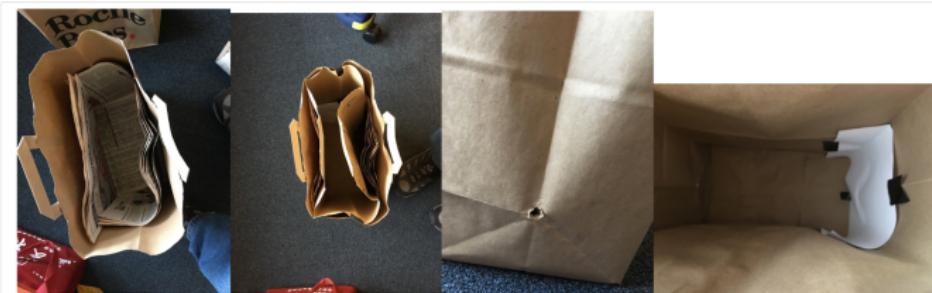


c) Difference image (b-a)    d) Crop upside down    e) True view

# Modèle de caméra : Sténopé

Quelques expériences intéressantes autour de ce modèle :

- Grocery bag pinhole camera
- <https://www.youtube.com/watch?v=-rhZaAM3F44>



# Modèle de caméra : Sténopé

Quelques expériences intéressantes autour de ce modèle :

- Grocery bag pinhole camera
- <https://www.youtube.com/watch?v=-rhZaAM3F44>



# Modèle de caméra : modèle géométrique

## Modèle du sténopé ou trou d'épingle (pin-hole)

- Modèle décrivant le processus de formation d'une image.
- Modélise une caméra par une projection perspective.

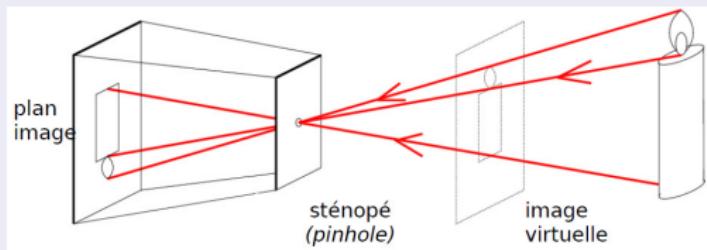


FIGURE – Source : A. Boucher

- Bon compromis entre la simplicité des équations associées et la proximité des phénomènes physiques modélisés.

# Modèle de caméra : modèle géométrique

## Modèle du sténopé

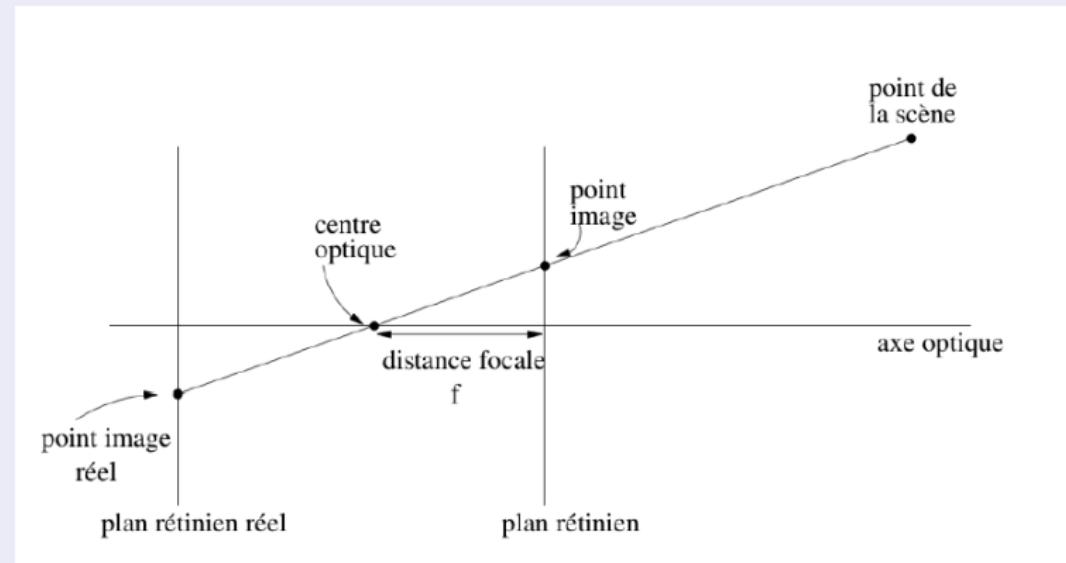


FIGURE – Source : E. Boyer, Irisa

# Modèle de caméra : modèle géométrique

## Modèle du sténopé

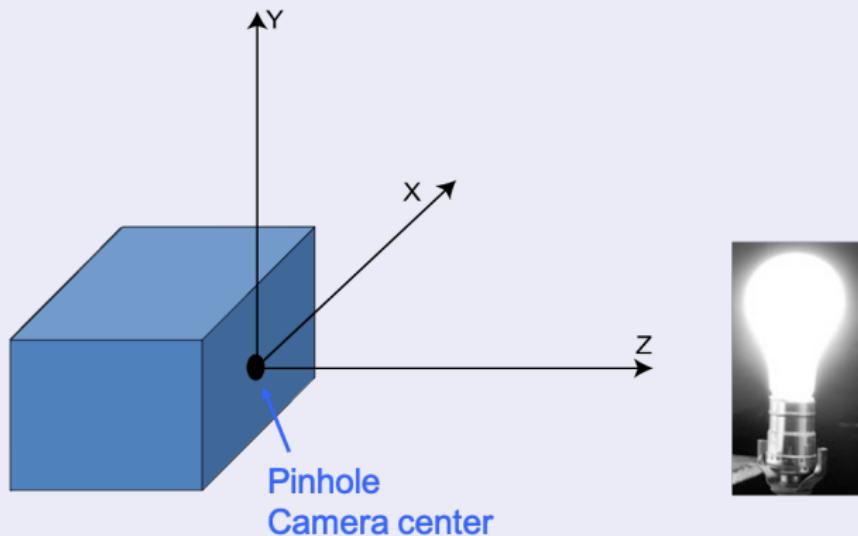


FIGURE – Source : MIT course

# Modèle de caméra : modèle géométrique

## Modèle du sténopé

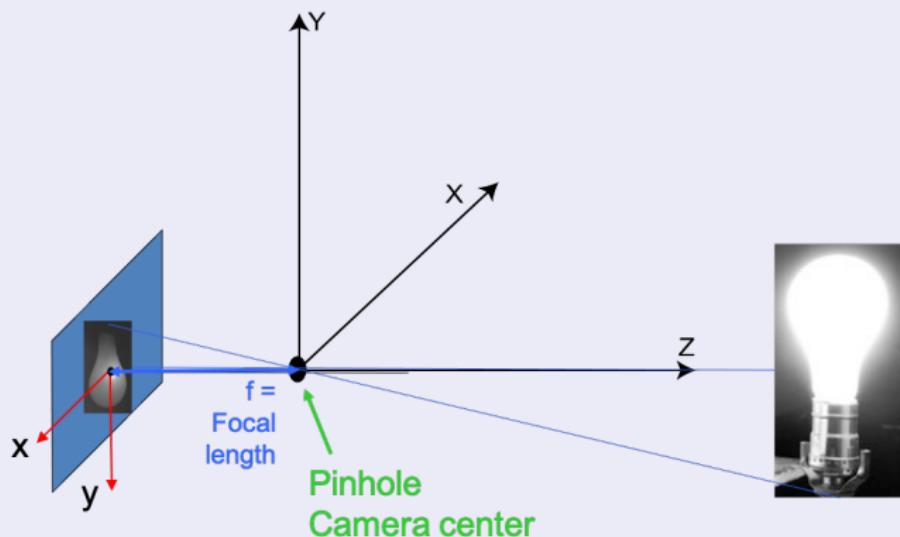


FIGURE – Source : MIT course

# Modèle de caméra : modèle géométrique

## Modèle du sténopé

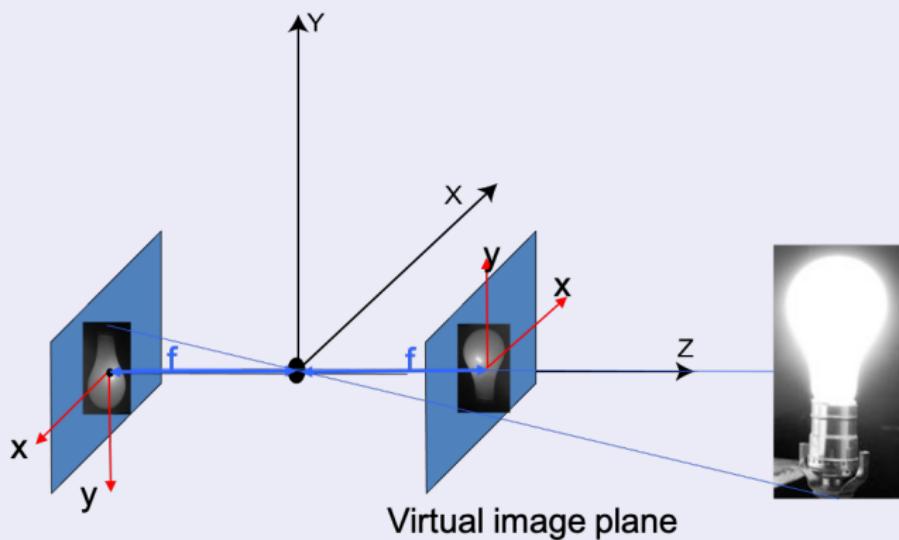


FIGURE – Source : MIT course

# Modèle de caméra : modèle géométrique

## Modèle du sténopé

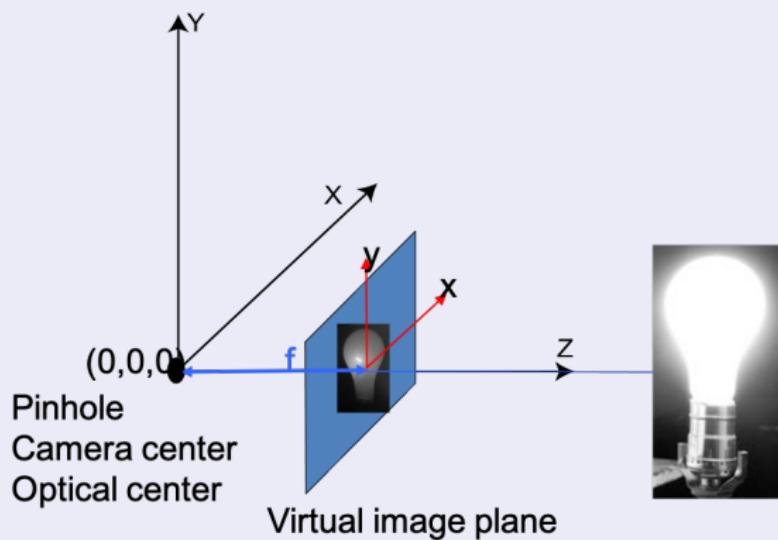
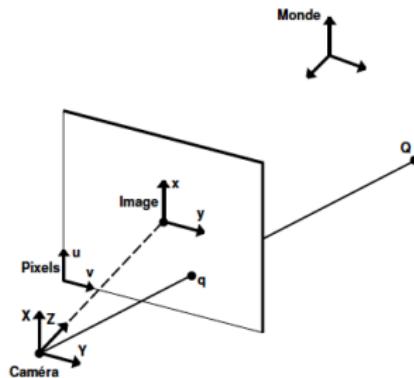


FIGURE – Source : MIT course

# Modèle de caméra : modèle algébrique

Dérivation des équations de projection à partir de 4 repères :

- le repère monde, 3D,
- un repère 3D attaché à la caméra (**repère caméra**),
  - ▶ Origine : centre de projection, axe Z : axe optique, X et Y parallèles au plan image.
- un repère 2D pour le plan image (**repère image**),
- le repère pixels (**repère pixel**).



**FIGURE – Source : P. Sturm**

# Modèle de caméra : modèle algébrique

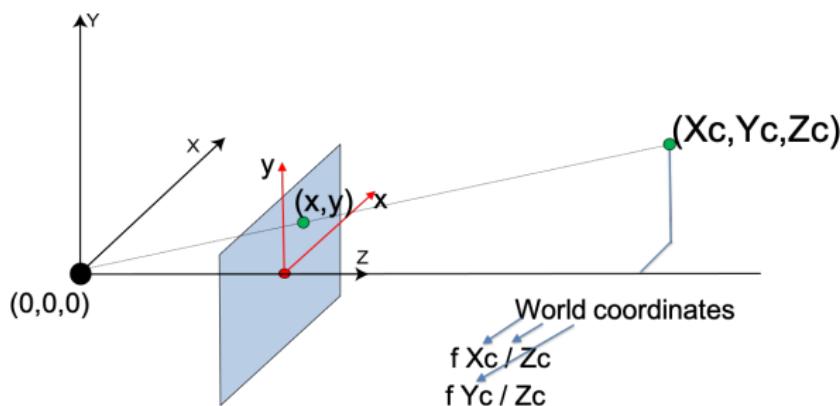
## En 3 étapes

- ① Prise en compte de la distance focale.
- ② Prise en compte des pixels.
- ③ Prise en compte des déplacements de la caméra.

# Modèle de caméra : modèle algébrique

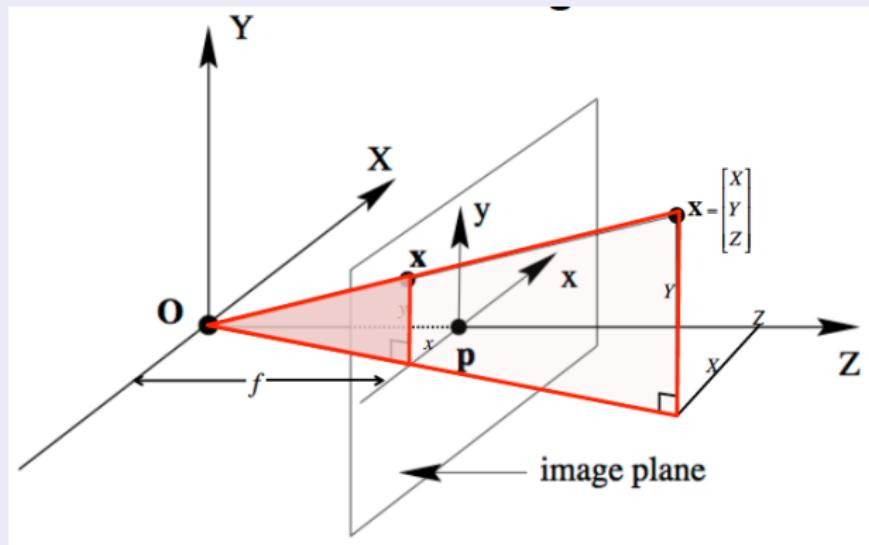
## Modèle du sténopé

- Une image : projection 2D du monde 3D.
- $P_c = (x_c, y_c, z_c)$  : point du monde 3D (repère caméra).
- $p = (x, y)$  : projection dans le plan image.
- $x = \frac{x_c \cdot f}{z_c}$ ,  $y = \frac{y_c \cdot f}{z_c}$  (relations entre triangles semblables).



# Modèle de caméra

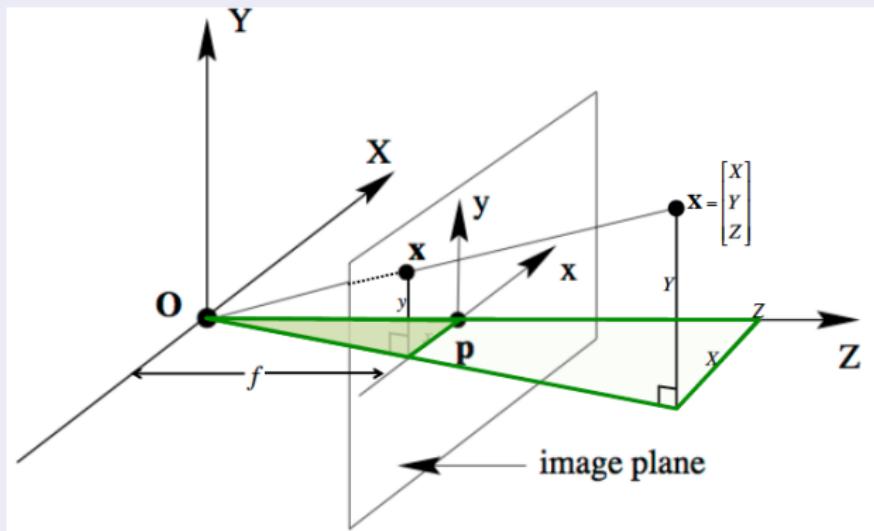
Modèle du sténopé : triangles similaires



Source : Xiao, Princeton

# Modèle de caméra

## Modèle du sténopé : triangles similaires



Source : Xiao, Princeton

# Modèle de caméra

## Coordonnées homogènes

- Un élément clé de la vision par ordinateur 3D.
- Permettent de définir une transformation entre deux référentiels avec une simple multiplication matricielle.
- Notation dans laquelle les vecteurs en  $n$  dimensions sont représentés par un vecteur en  $n + 1$  dimension.
- Exemple :  $P = \begin{pmatrix} x \\ y \end{pmatrix}$  devient  $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$
- Invariance à la multiplication par une constante<sup>a</sup>.
- Conversion à partir de coordonnées homogènes :

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow \begin{pmatrix} \frac{x}{w} \\ \frac{y}{w} \end{pmatrix}$$

# Parenthèse : coordonnées homogènes

## Géométrie projective : définition

Géométrie a priori non métrique qui a comme objectif d'étudier les propriétés des configurations d'ensembles de droites et de points affectés par des projections.

Eléments géométriques de base :

- Points
- Lignes droites
- Coniques
- ...

Les coordonnées homogènes vont permettre de rendre les calculs possibles dans l'espace projectif (comme les coordonnées cartésiennes dans l'espace euclidien) et de représenter sous forme matricielle l'ensemble des transformations.

# Parenthèse : coordonnées homogènes

## Point : Passage de l'espace euclidien à l'espace projectif

- Un point  $P$  d'un espace projectif est repéré par ses coordonnées homogènes.

- Chaque point  $P' = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$  de l'espace euclidien,  $P' \in \mathbb{R}^3$  est associé à

$$P = \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, P \in \mathbb{P}^3$$
 de l'espace projectif.

- $\mathbb{P}^3 = \mathbb{R}^3 - (0, 0, 0)$  est l'espace projectif 3D.
- Les coordonnées homogènes d'un point dans un espace projectif ne sont pas uniques :

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \text{ et } \begin{pmatrix} wx \\ wy \\ wz \\ w \end{pmatrix} \text{ représentent le même point } (w \neq 0).$$

# Parenthèse : coordonnées homogènes

## Point : Passage de l'espace projectif à l'espace euclidien

- Un point de l'espace projectif dont la dernière coordonnée homogène est non nulle peut être associé à un point de l'espace euclidien :

$$P = \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}, P \in \mathbb{P}^3 \text{ est associé à } P' = \begin{pmatrix} \frac{x}{w} \\ \frac{y}{w} \\ \frac{z}{w} \\ \frac{1}{w} \end{pmatrix}, P' \in \mathbb{R}^3$$

- Un point de l'espace projectif dont la dernière composante est nulle est appelé **point à l'infini** (il n'a pas d'équivalent dans l'espace euclidien)
- Vecteur homogène et vecteur non-homogène

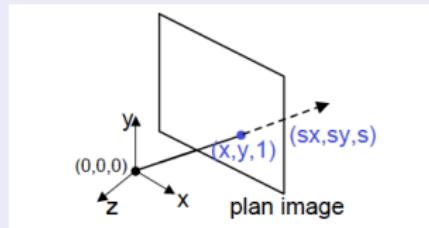
$$P = (x, y, z, w) = w(x', y', z', 1) = w\bar{P}$$

avec  $\bar{P}$  le vecteur augmenté.

# Parenthèse : coordonnées homogènes

## Pourquoi les coordonnées homogènes ?

- Permettent de représenter les points à l'infini, les homographies, la projection perspective, les relations multi-vues.
- Intuition géométrique :
  - ▶ Un point de l'image est un rayon dans l'espace projectif.
  - ▶ Chaque point  $(x, y)$  du plan est représenté par un rayon  $(sx, sy, s)$ .
  - ▶ Un point de  $\mathbb{P}^2$  est une direction de  $\mathbb{R}^3$ .



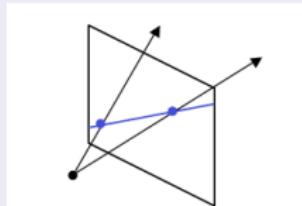
- ▶ Tous les points du rayon sont équivalents :

$$(x, y, 1) \approx (sx, sy, s)$$

# Parenthèse : coordonnées homogènes

## Droites projectives

Qu'est-ce qu'une droite dans l'espace projectif ?



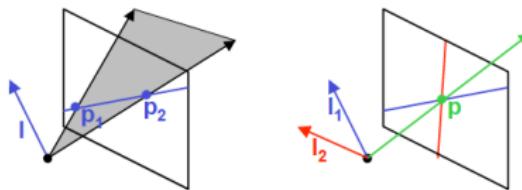
- Une droite est un plan formé de rayons passant par l'origine.
- Tous les rayons  $(x, y, z)$  satisfont :  $ax + by + cz = 0$ , ce qui donne en notation vectorielle :

$$0 = [a \ b \ c] \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$0 = I^T p = p^T I \text{ (dualité point - droite) avec } I^T = (a, b, c) \text{ et } p^T = (x, y, z)$$

- Une droite est représentée par un vecteur de dimension 3,  $I$ , en coordonnées homogènes.

# Parenthèse : coordonnées homogènes



## Principe de dualité point-droite

Une droite  $\mathbf{l}$  est un vecteur à 3 coordonnées. Elle est perpendiculaire à tout point (rayons) sur la droite  $\mathbf{l}^T \mathbf{p} = 0$

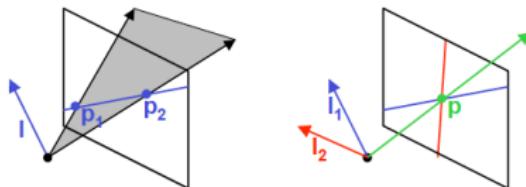
- Quelle est la droite  $\mathbf{l}$  passant par  $\mathbf{p}_1$  et  $\mathbf{p}_2$  ?  
 $\mathbf{l}$  normale au plan formé par les deux rayons.

$$\mathbf{l} = \mathbf{p}_1 \wedge \mathbf{p}_2 \text{ (produit vectoriel)}$$

- Quelle est l'intersection de deux droites  $\mathbf{l}_1$  et  $\mathbf{l}_2$  ?  
Le rayon  $\mathbf{p}$  est perpendiculaire à  $\mathbf{l}_1$  et  $\mathbf{l}_2$

$$\mathbf{p} = \mathbf{l}_1 \wedge \mathbf{l}_2 \text{ (produit vectoriel)}$$

# Parenthèse : coordonnées homogènes

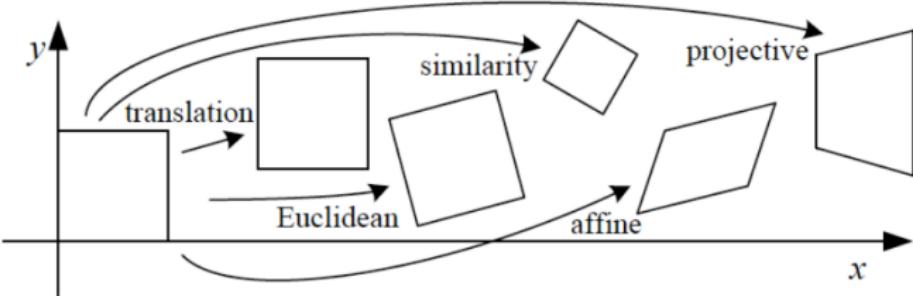


## Principe de dualité point-droite

A chaque théorème de géométrie projective 2D, correspond un théorème dual, construit à partir de l'original, en inter-changeant le rôle des points et des droites.

# Parenthèse : coordonnées homogènes

## Transformations 2D



Example: translation

$$x' = x + t \quad x' = [ \mathbf{I} \quad t ] \bar{x} \quad \bar{x}' = \begin{bmatrix} \mathbf{I} & t \\ 0^T & 1 \end{bmatrix} \bar{x}$$

$$\begin{bmatrix} \text{teal} \\ \text{light green} \end{bmatrix} = \begin{bmatrix} \text{teal} \\ \text{light green} \end{bmatrix} + \begin{bmatrix} \text{tx} \\ \text{ty} \end{bmatrix}$$

$$\begin{bmatrix} \text{teal} \\ \text{light green} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \text{tx} \\ 0 & 1 & \text{ty} \end{bmatrix} \cdot \begin{bmatrix} \text{teal} \\ \text{light green} \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \text{teal} \\ \text{light green} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \text{tx} \\ 0 & 1 & \text{ty} \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \text{teal} \\ \text{light green} \\ 1 \end{bmatrix}$$

Source : A. Torralba - Ici  $\mathbf{I}$  est la matrice identité.

# Parenthèse : coordonnées homogènes

## Translation et Rotation

- Transformation euclidienne.

$$\bar{x}' = [\mathbf{R} \ t]\bar{x}$$

avec

$$\mathbf{R} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

la matrice de rotation telle que  $\mathbf{R}\mathbf{R}^T = \mathbf{I}$  et  $|\mathbf{R}| = 1$ .

# Parenthèse : coordonnées homogènes

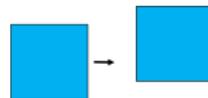
## Transformations affines

- Transformation affine

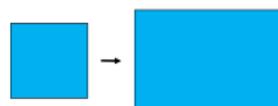
$$\mathbf{x}' = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix} \bar{\mathbf{x}}$$

# Parenthèse : coordonnées homogènes

Translation



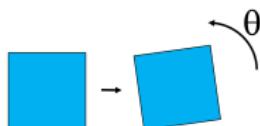
Scaling



$$\begin{matrix} x' \\ y' \\ 1 \end{matrix} = \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{matrix} x \\ y \\ 1 \end{matrix}$$

$$\begin{matrix} x' \\ y' \\ 1 \end{matrix} = \begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{matrix} x \\ y \\ 1 \end{matrix}$$

Rotation



$$\begin{matrix} x' \\ y' \\ 1 \end{matrix} = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{matrix} x \\ y \\ 1 \end{matrix}$$

Shearing



$$\begin{matrix} x' \\ y' \\ 1 \end{matrix} = \begin{pmatrix} 1 & a_x & 0 \\ a_y & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{matrix} x \\ y \\ 1 \end{matrix}$$

# Parenthèse : coordonnées homogènes

## Homographies 2D

Transformation linéaire entre deux plans projectifs  $\pi_1$  et  $\pi_2$

$$\mathbf{q} = \mathcal{H}\mathbf{p} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \mathbf{p} \text{ avec } \mathbf{p} \in \pi_1, \mathbf{q} \in \pi_2$$

- Plans projectifs identiques : transformation
- Plans projectifs non identiques : transfert d'un plan projectif vers l'autre
- L'homographie préserve l'alignement ( une droite d'un plan  $\pi_1$  est transformée en une droite dans l'espace  $\pi_2$  mais les angles et les distances ne sont pas préservées.

# Parenthèse : coordonnées homogènes

## Homographies 2D



image initiale dans  $\pi_1$



image transformée dans  $\pi_2$

Source : F. Cabestaing

# Parenthèse : coordonnées homogènes

Transformations : tableau de synthèse.

Transformation	Matrix	# DoF	Preserves	Icon
translation	$\begin{bmatrix} \mathbf{I} & \mathbf{t} \end{bmatrix}_{2 \times 3}$	2	orientation	
rigid (Euclidean)	$\begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix}_{2 \times 3}$	3	lengths	
similarity	$\begin{bmatrix} \mathbf{sR} & \mathbf{t} \end{bmatrix}_{2 \times 3}$	4	angles	
affine	$\begin{bmatrix} \mathbf{A} \end{bmatrix}_{2 \times 3}$	6	parallelism	
projective	$\begin{bmatrix} \tilde{\mathbf{H}} \end{bmatrix}_{3 \times 3}$	8	straight lines	

**Table 2.1** Hierarchy of 2D coordinate transformations, listing the transformation name, its matrix form, the number of degrees of freedom, what geometric properties it preserves, and a mnemonic icon. Each transformation also preserves the properties listed in the rows below it, i.e., similarity preserves not only angles but also parallelism and straight lines. The  $2 \times 3$  matrices are extended with a third  $[0^T \ 1]$  row to form a full  $3 \times 3$  matrix for homogeneous coordinate transformations.

# Modèle de caméra : modélisation algébrique

Fin de la parenthèse, retour au modèle de caméra

Dérivation des équations de projection à partir de 4 repères :

- le repère monde, 3D,
- un repère 3D attaché à la caméra (**repère caméra**),
  - ▶ Origine : centre de projection, axe Z : axe optique, X et Y parallèles au plan image.
- un repère 2D pour le plan image (**repère image**),
- le repère pixels (**repère pixel**).

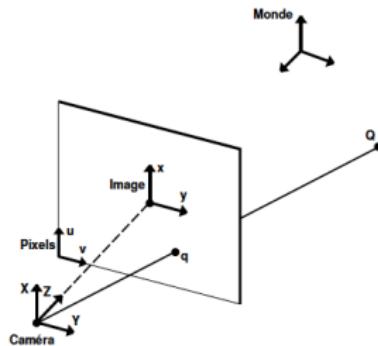


FIGURE – Source : P. Sturm

# Modèle de caméra : modélisation algébrique

Fin de la parenthèse, retour au modèle de caméra

On modélise la projection perspective (point dans le repère caméra) par une transformation projective (centre optique à l'origine du repère, plan image  $\pi$  à une distance  $f$  du centre optique, plan image orthogonal à l'axe  $z$ .)

## Modèle du sténopé

- Projection perspective

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{x_c \cdot f}{z_c} \\ \frac{y_c \cdot f}{z_c} \end{bmatrix}$$

- En coordonnées homogènes

$$\begin{bmatrix} x.w \\ y.w \\ w \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{bmatrix} \cdot \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}$$

# Modèle de caméra

## Modèle du sténopé : prise en compte des pixels (repère pixel)

- $p = (x, y)$  : projection dans le plan image.
- $p_i = (x_i, y_i)$  : coordonnées des pixels.
- Prise en compte de l'échantillonnage, d'une éventuelle distorsion radiale, ...
- Le changement de repère (image à pixels) nécessite une translation et un changement d'unité, i.e. repère image = repère métrique alors que dans le repère pixels l'unité est le nombre de pixels

Soit

- ▶  $(x_0, y_0)$  : coordonnées du coin en bas à gauche de l'image dans le repère image.
- ▶  $(k_x, k_y)$  : facteurs d'échelle en pixels par mm par exemple, i.e. densité de pixels en direction de l'axe des u et des v (repère pixel) (par exemple en nb pixels par mm).
- En coordonnées homogènes :

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

# Modèle de caméra

## Modèle du sténopé : prise en compte des pixels (repère pixel)

Modélisation de la projection complète du repère caméra vers le repère pixels.

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \sim \begin{bmatrix} k_x f & 0 & k_x x_0 & 0 \\ 0 & k_y f & k_y y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}$$

# Modèle de caméra

## Modèle du sténopé : prise en compte des déplacements

- Les points 3D ont été représentés dans le repère caméra.
- Nécessité de prendre un repère statique attaché à la scène 3D pour prendre en compte les déplacements que les caméras peuvent effectuer.
- Entrée en jeu du **repère monde** ou repère environnement.
- La position des points 3D et des caméras sont décrites par rapport à ce repère.
- c.f. calibration

# Plan

- 1 Introduction
- 2 Projection perspective
- 3 Modèle de caméra
- 4 Retour sur la projection perspective
- 5 Calibration d'une caméra
- 6 Triangulation
- 7 Vision stéréoscopique
- 8 Structure from Motion
- 9 Conclusion

# Projection perspective

La projection peut être trompeuse



# Projection perspective

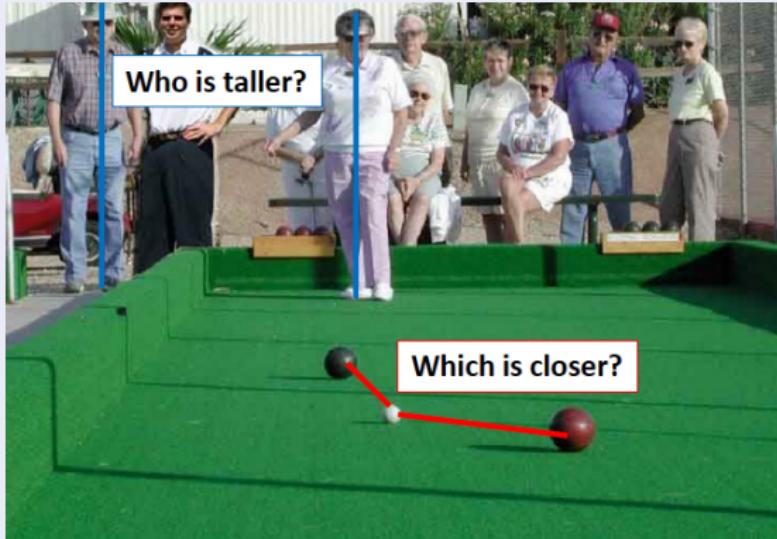
La projection peut être trompeuse



# Projection perspective

Effet : perte de la taille

La taille des objets n'est pas préservée.



Source : J. Hays.

# Projection perspective

## Effets

Effet de la perspective : plus les objets sont éloignés et plus ils sont petits dans l'image.

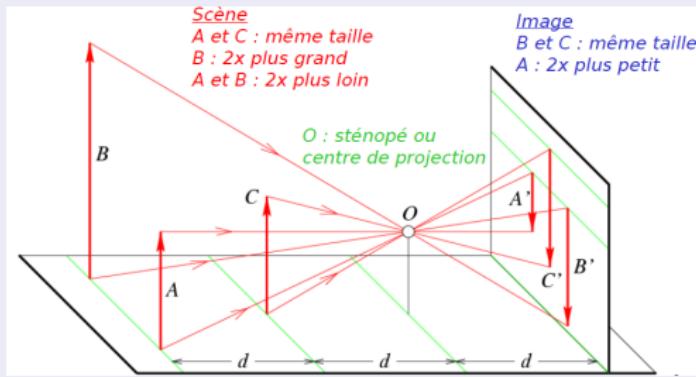


FIGURE – Source : A. Boucher

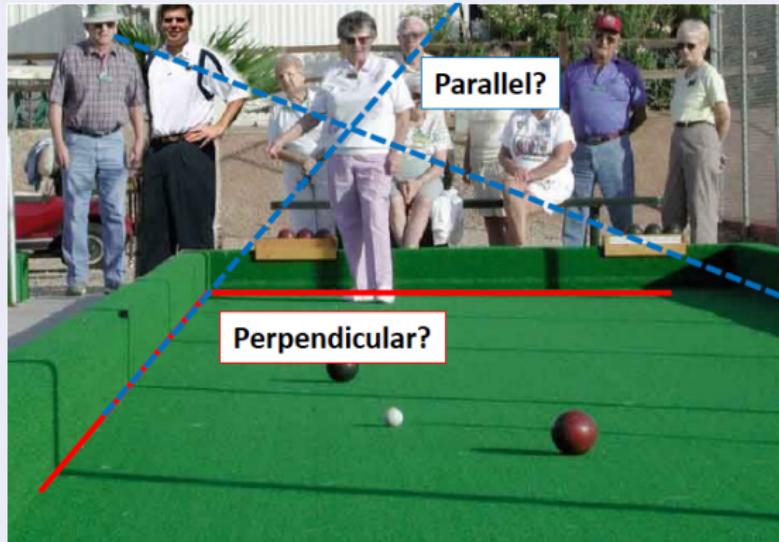
Les proportions sont respectées :

- même distance : taille image proportionnelle à taille objet.
- même taille : taille image inversement proportionnelle à la distance.

# Projection perspective

Effet : perte des angles

Les angles ne sont pas préservés

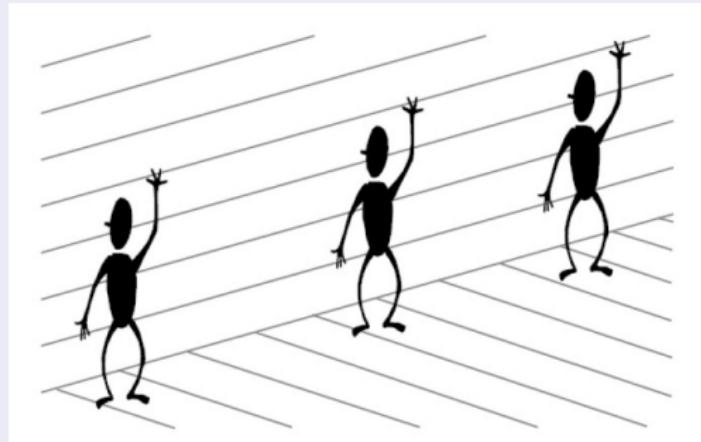


Source : J. Hays.

# Projection perspective

## Indices visuels de perspective

- Perception des indices de perspective dans une image par des connaissances a priori sur les objets.
- Perception humaine tient compte des règles de perception perspective.
- Si non respect des proportions, la vision se trompe sur la taille ou la distance.



# Projection perspective

## Indices de perspective

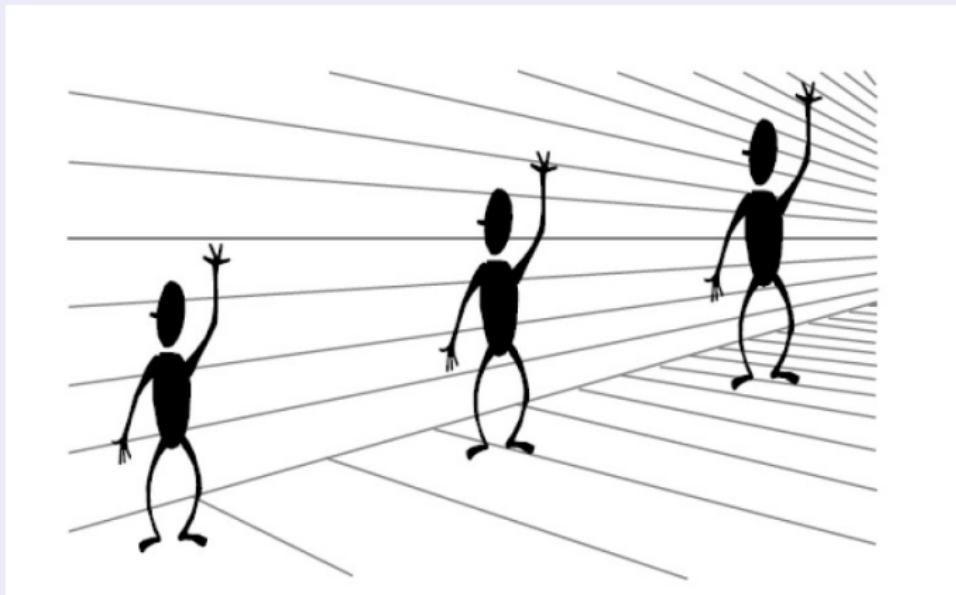


FIGURE – Source :F. Devernay

# Projection perspective

## Indices de perspective

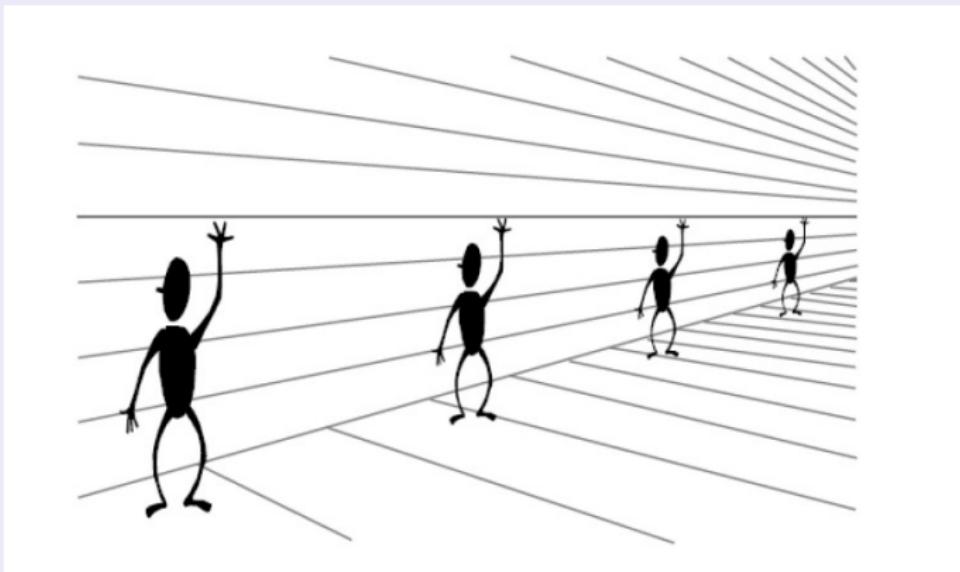


FIGURE – Source :F. Devernay

# Projection perspective : illusions d'optique



FIGURE – Source :F. Devernay

- Toutes les images de voiture ont la même taille.
- L'effet de perspective nous trompe sur leur taille en tant qu'objets de la scène.

# Projection perspective

## Indices de perspective

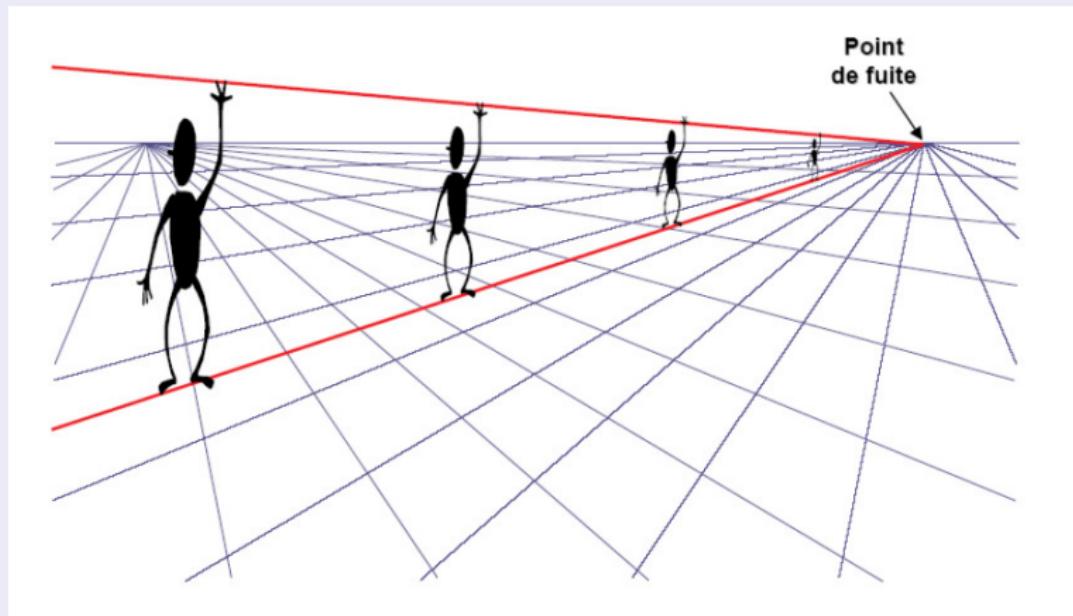


FIGURE – Source :F. Devernay

# Projection perspective

## Point de fuite

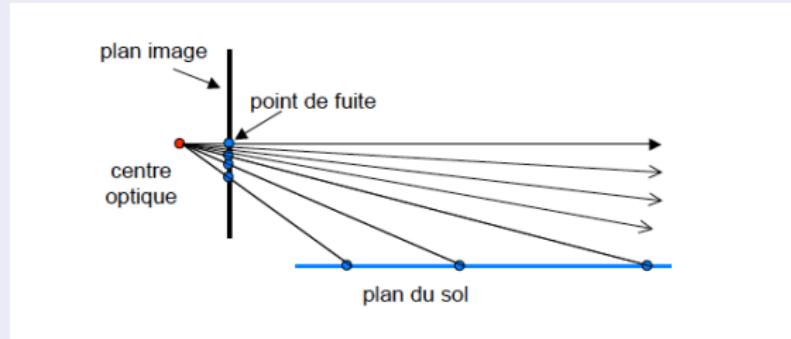


FIGURE – Source :F. Devernay

Point de fuite :

- Projection d'un point à l'infini.

# Projection perspective

## Point de fuite (2D)

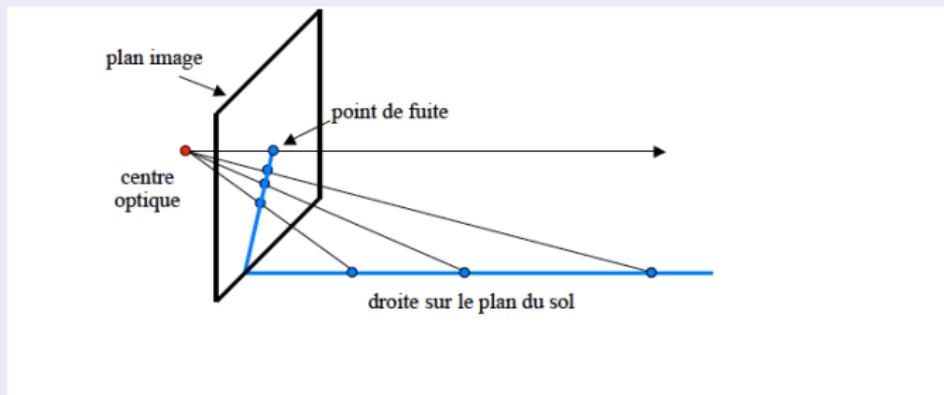


FIGURE – Source :F. Devernay

# Projection perspective

## Points de fuite

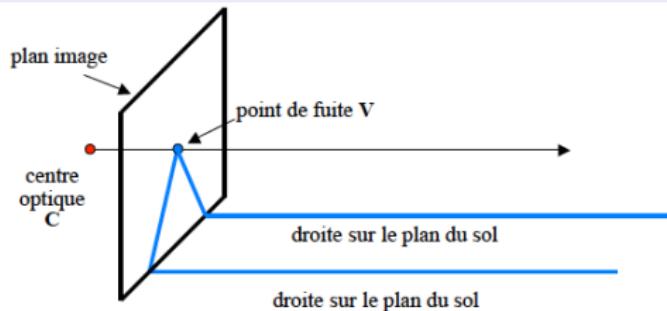


FIGURE – Source :F. Devernay

Propriétés :

- Deux droites parallèles ont le même point de fuite.
- Le rayon issu de  $C$  passant par  $v$  est parallèle aux droites.
- Une image peut avoir plus d'un point de fuite.

# Projection perspective

## Point de fuite et ligne d'horizon

- Les lignes parallèles se rejoignent en un point de l'image (point de fuite)
  - ▶ Parallèle = origine différente - même direction.
  - ▶ Le point de fuite vaut pour cette direction
- L'union de tous les points de fuite est la ligne d'horizon (ou ligne de fuite).

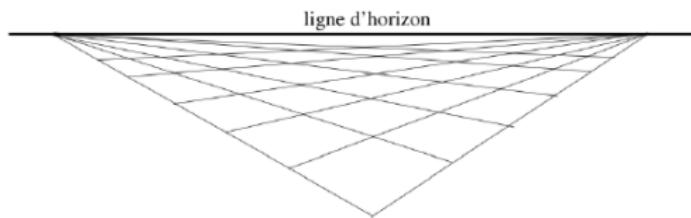
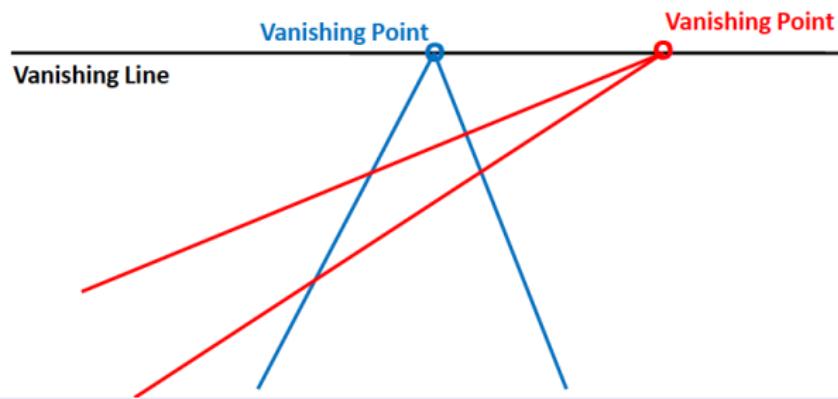


FIGURE – Source :E. Boyer

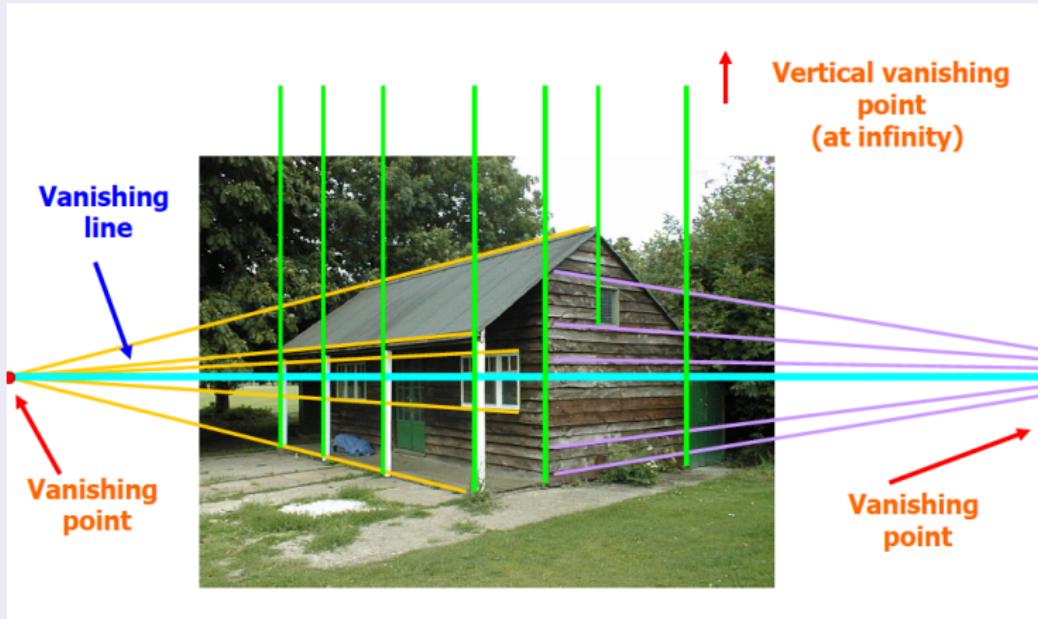
# Projection perspective

## Point de fuite et ligne d'horizon



# Projection perspective

## Point de fuite et ligne d'horizon



# Plan

- 1 Introduction
- 2 Projection perspective
- 3 Modèle de caméra
- 4 Retour sur la projection perspective
- 5 Calibration d'une caméra
- 6 Triangulation
- 7 Vision stéréoscopique
- 8 Structure from Motion
- 9 Conclusion

# Calibration d'une caméra

## Définition

La calibration (ou etalonnage) de la caméra consiste à déterminer les paramètres entre les coordonnées du monde et les coordonnées image (et vice-versa)

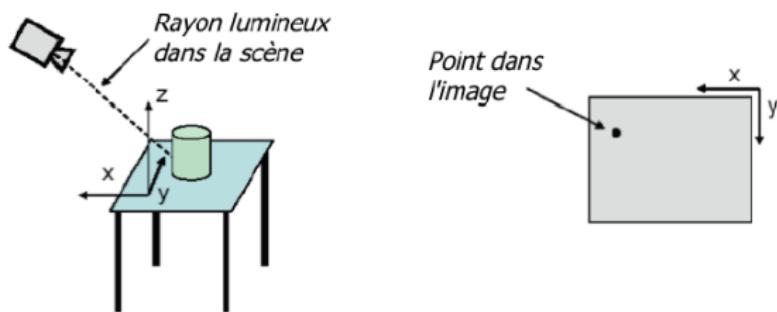


FIGURE – Source :A. Boucher

# Calibration d'une caméra

- Le modèle de la caméra décrit la transformation du système de coordonnées de la scène vers celui de l'image.

scène → caméra → image

- paramètres extrinsèques

- ▶ Relation entre le référentiel environnement et le référentiel caméra.
- ▶ Position et orientation de la caméra par rapport à la scène.

- paramètres intrinsèques

- ▶ Propriétés optiques et géométriques de la caméra.
- ▶ Ne changent pas si on déplace la caméra.

# Calibration d'une caméra

## Trois (quatre) systèmes de coordonnées distincts

- L'environnement (ou le monde de la scène)  $R_e$ 
  - ▶ système de référence connu. Tout est défini (caméra, objet, ...) par rapport à ce système de coordonnées.
- La camera  $R_c$ 
  - ▶ Système dans lequel s'effectue la projection des points 3D vers 2D.
- L'image  $R_i$ 
  - ▶ Système établissant la référence aux pixels.

$$P_e(x_e, y_e, z_e) \rightarrow P_c(x_c, y_c, z_c) \rightarrow P_i(x_i, y_i)$$

# Calibration d'une caméra

Trois systèmes de coordonnées distincts

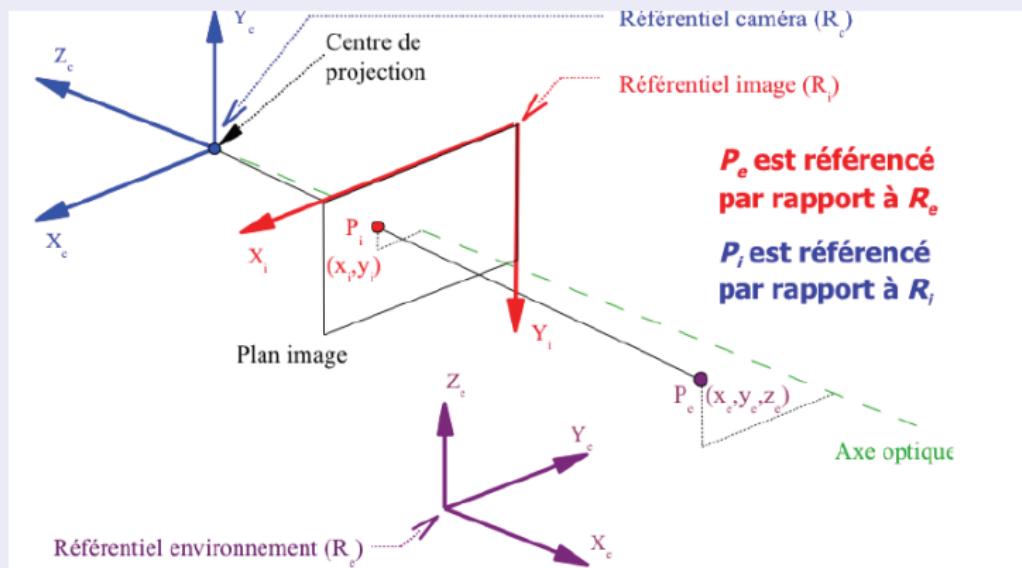


FIGURE – Source :A. Boucher

# Calibration d'une caméra

## Paramètres extrinsèques

- La position et l'orientation : transformation rigide 3D
  - ▶ Translation  $T$  pour faire correspondre les origines des systèmes de coordonnées.
  - ▶ Rotation  $R$  pour faire correspondre les axes des systèmes.

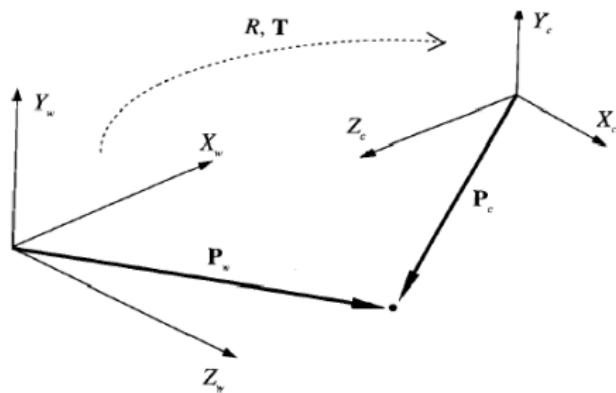


FIGURE – Source :A. Boucher

# Calibration d'une caméra

## Paramètres extrinsèques

Transformation rigide décrite par les paramètres extrinsèques.

$$P_c = RP_e + T$$

- Ecriture matricielle

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} R \\ & T \end{bmatrix} \cdot \begin{bmatrix} x_e \\ y_e \\ z_e \end{bmatrix} + \begin{bmatrix} T \end{bmatrix}$$

- Coordonnées homogènes

$$\begin{bmatrix} x_c \cdot w \\ y_c \cdot w \\ z_c \cdot w \\ w \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_e \\ y_e \\ z_e \\ 1 \end{bmatrix}$$

# Calibration d'une caméra

## Du monde à l'image

- On a (paramètres internes)

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & x_0 \\ 0 & k_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

- Donc

$$\begin{bmatrix} x_i \cdot w \\ y_i \cdot w \\ w \end{bmatrix} = \begin{bmatrix} k_x & 0 & x_0 \\ 0 & k_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \cdot w \\ y \cdot w \\ w \end{bmatrix}$$

- Et

$$\begin{bmatrix} x_i \cdot w \\ y_i \cdot w \\ w \end{bmatrix} = \begin{bmatrix} k_x & 0 & x_0 \\ 0 & k_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/f \end{bmatrix} \cdot \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}$$

# Calibration d'une caméra

## Du monde à l'image

- On en déduit

$$\begin{bmatrix} x_i \cdot w \\ y_i \cdot w \\ w \end{bmatrix} = \begin{bmatrix} k_x & 0 & u_0 \\ 0 & k_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{bmatrix} \cdot \begin{bmatrix} R & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_e \\ y_e \\ z_e \\ 1 \end{bmatrix}$$

$$P_i = M \cdot P_e$$

- $M$  : matrice  $3 \times 4$  contenant les paramètres intrinsèques et extrinsèques : il s'agit de la **matrice de projection**.
- Tous les paramètres sont définis à un facteur d'échelle près (on ne connaît pas la profondeur des points).
- 11 inconnues.

# Calibration d'une caméra

$$\begin{pmatrix} x_i \cdot w \\ y_i \cdot w \\ w \end{pmatrix} = \begin{pmatrix} \text{Transformation} \\ \text{paramètres} \\ \text{intrinsèques} \end{pmatrix} \xrightarrow{\text{(Projection)}} \begin{pmatrix} \text{Transformation} \\ \text{paramètres} \\ \text{extrinsèques} \end{pmatrix} \begin{pmatrix} x_e \\ y_e \\ z \\ 1 \end{pmatrix}$$

Matrice  $3 \times 3$       Matrice  $4 \times 3$       Matrice  $4 \times 4$

$$P_i = M \cdot P_e = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \cdot P_e$$

FIGURE – Source :A. Boucher

Comme les paramètres sont définis à un facteur d'échelle près, on a  $m_{34} = 1$ .

# Calibration d'une caméra

## Notation classique

$$\mathbf{K}_{\text{intrinsics}} \left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right] \left[ \begin{array}{ccc} \mathbf{R} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{array} \right] \left[ \begin{array}{cc} \mathbf{I}_{3 \times 3} & -\mathbf{c} \\ 0 & 0 & 0 & 1 \end{array} \right]$$

projection                      rotation                      translation

$\mathbf{K} \left[ \begin{array}{c|c} \mathbf{R} & -\mathbf{R}\mathbf{c} \end{array} \right]$

$\mathbf{t}$

# Calibration d'une caméra

## Paramètres extrinsèques et intrinsèques

- **Paramètres extrinsèques** : matrice de rotation  $R$  et la position  $c$  de la caméra par rapport au monde extérieur.
- **Paramètres intrinsèques** : caractéristiques internes de la caméra : paramètres  $f$ ,  $k_x$ ,  $k_y$ ,  $x_0$  et  $y_0$  et regroupés dans une matrice triangulaire supérieure  $K$ .

$$K = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}.$$

avec  $\alpha_u = k_x f$  et  $\alpha_v = k_y f$  exprimant la distance focale en nombre de pixels et  $u_0 = k_x x_0$  et  $v_0 = k_y y_0$  les coordonnées du point principal dans le repère pixel.

$K$  est la **matrice de calibrage** de la caméra.

# Calibration d'une caméra

- Le processus de détermination de  $K$  s'appelle la calibration de la caméra.

## Calibration

Retrouver les paramètres internes de la caméra (mais souvent aussi les paramètres externes).

## Calcul de pose

Calculer les paramètres externes. On suppose souvent que les paramètres internes sont connus.

# Calibration d'une caméra

## Différentes approches

- Avec un objet 3D connu : ses points sont connus précisément et sont facilement identifiables dans les images.
- Avec un objet plan dont la géométrie est connue : les paramètres sont calculés à partir de l'observation de plusieurs vues de l'objet sans avoir besoin de connaître leur position (**Facilité de mise en œuvre**).
- Auto-calibration : à partir de plusieurs images d'un objet inconnu , on cherche à calculer les positions de la caméra et les paramètres intrinsèques (mouvement de la caméra).

## Grandes étapes

- Identifier la projection des éléments dans les images.
- Minimiser une fonction des paramètres et des entités.

# Calibration d'une caméra

## Méthode directe

- On suppose que l'on connaît :
  - ▶ Les positions  $M_i$  de  $n$  points dans la scène
  - ▶ Les positions  $m_i$  de leurs projections dans le plan image
- Il faut calculer les paramètres qui minimisent le résidu :

$$\sum_{i=1}^n (m_i - \mathbf{M}M_i)^2$$

- Il faut suffisamment de **bons** points.
- Pas de solution analytique, donc méthodes numériques itératives.

# Calibration d'une caméra avec un objet 3D

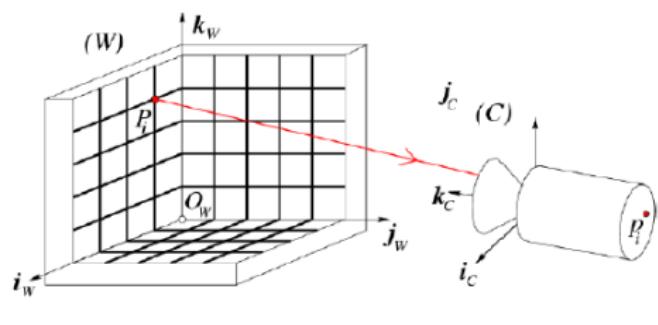
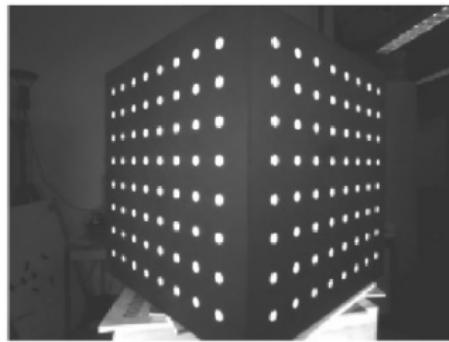
- Processus standard :

- ▶ Prendre une image d'un objet dont la structure est parfaitement connue : **mire de calibrage** :
  - ★ Correspondances : points / lignes 3D avec leur image 2D
  - ★ Les points 3D doivent être en position générale (non-planaire)
- ▶ Repère attaché à la mire : les cibles sont connues avec une très haute précision.
- ▶ Dans l'image, les cibles doivent être extraites par traitement.
- ▶ Mises en correspondance : pour chaque cible dans l'image, il faut déterminer la cible 3D dont elle est la projection.
- ▶ Processus semi-automatique.

# Calibration d'une caméra : estimation de M

## Mire de calibration

- Utilisation d'une mire pour calibrer les caméras.
- La mire doit contenir des formes faciles à reconnaître et dont la position est très bien connue selon le référentiel de l'environnement.
- On mesure le centre des points blancs qui sont connus à l'avance et facile à identifier.
- La cible doit comporter des éléments non-coplanaires.



# Calibration d'une caméra

## Avec une mire de calibration

Première étape : correspondance entre points 3D et points 2D

- Trouver des points saillants.
- Trouver des bordures des carrés et trouver des lignes.
- Intersection de ces lignes pour avoir des coordonnées sous-pixels.
- Retrouver la configuration des points (i.e. leurs coordonnées)

# Calibration d'une caméra

Première étape : correspondance entre points 3D et points 2D

## Exemple

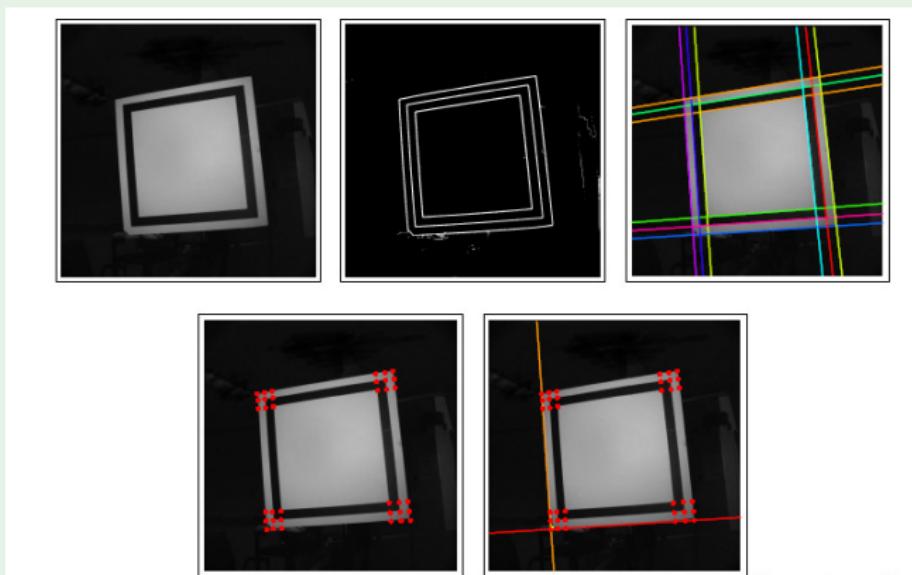


FIGURE – Source : S. Roy

# Calibration d'une caméra

Exemples de mire de calibration

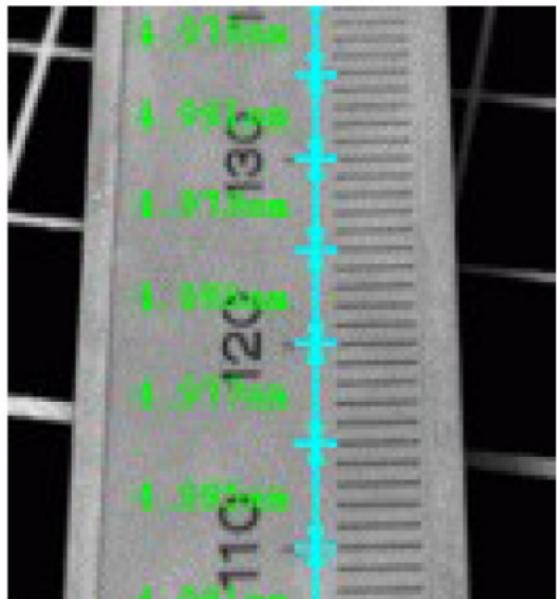
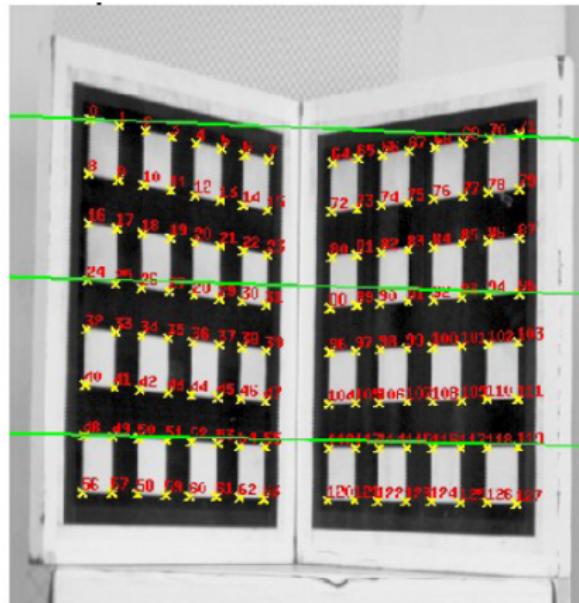


FIGURE – Source :A. Boucher

# Calibration d'une caméra

Deuxième étape : estimation de  $M$ .

- Processus de détermination de  $M$ .
- Il y a 11 inconnues donc prendre un nombre de points suffisants pour la résolution.
- Résolution par des méthodes numériques de minimisation d'erreur (decomposition SVD, méthode des moindres carrés,...)
- Implémentation semi-automatique dans OpenCV.

[http://docs.opencv.org/doc/tutorials/calib3d/camera\\_calibration/camera\\_calibration.html](http://docs.opencv.org/doc/tutorials/calib3d/camera_calibration/camera_calibration.html)

[http://docs.opencv.org/modules/calib3d/doc/camera\\_calibration\\_and\\_3d\\_reconstruction.html](http://docs.opencv.org/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html)

# Calibration d'une caméra avec une mire 3D

Deuxième étape : estimation de  $\mathbf{M}$ .

## Données

Les  $n$  points de référence 3D :  $M_i = \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix}$  et leur projection dans le plan image  
 $m_i = \begin{pmatrix} u_i \\ v_i \end{pmatrix}$

## But

Déterminer  $\mathbf{M}$ , telle que  $m_i = \mathbf{M}M_i$

# Calibration d'une caméra avec une mire 3D

Deuxième étape : estimation de  $M$ .

$$\begin{bmatrix} su_i \\ sv_i \\ s \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$

$$① \quad su_i = m_{11}X_i + m_{12}Y_i + m_{13}Z_i + m_{14}$$

$$② \quad sv_i = m_{21}X_i + m_{22}Y_i + m_{23}Z_i + m_{24}$$

$$③ \quad s = m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34}$$

$$(1) - (3)u_i \rightarrow (m_{11} - u_i m_{31})X_i + (m_{12} - u_i m_{32})Y_i + (m_{13} - u_i m_{33})Z_i + (m_{14} - u_i m_{34}) = 0$$

$$(2) - (3)v_i \rightarrow (m_{21} - v_i m_{31})X_i + (m_{22} - v_i m_{32})Y_i + (m_{23} - v_i m_{33})Z_i + (m_{24} - v_i m_{34}) = 0$$

2n équations linéaires à 12 inconnues.

Calibration d'une caméra avec une mire 3D

$$\begin{bmatrix} & & & & \cdot & & & \\ & & & & \cdot & & & \\ & & & & \cdot & & & \\ X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & -u_i X_i & -u_i Y_i & -u_i Z_i & -u_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_i X_i & -v_i Y_i & -v_i Z_i & -v_i \\ & & & & \cdot & & & \\ & & & & \cdot & & & \\ & & & & \cdot & & & \\ \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ \vdots \\ m_{32} \\ m_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

L'indice  $i$  représente le  $i$ -ème appariement entre points 3D et 2D détectés dans l'image (6 appariements sont nécessaires au minimum).

### Résolution aux moindres carrés ( $Gp = 0$ )

D'autres méthodes de résolution non linéaires : Ceres-Solver de Google (<http://ceres-solver.org/>)

# Calibration d'une caméra : A RETENIR

La seule chose à finalement retenir

$$x = K[R|t]X$$

# A RETENIR

Avec la même caméra mais deux points de vues

$$x_1 = K[R_1|t_1]X$$

$$x_2 = K[R_2|t_2]X$$

Vision stéréoscopique

# A RETENIR

Avec la même caméra mais différents points de vues

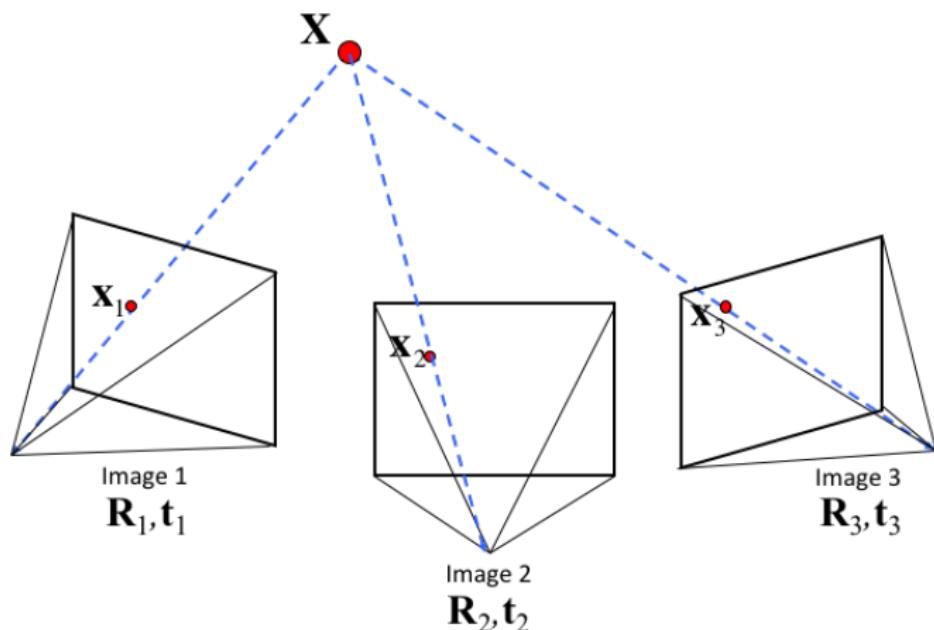
$$x_1 = K[R_1|t_1]X$$

$$x_2 = K[R_2|t_2]X$$

$$x_3 = K[R_3|t_3]X$$

## Vision multi-vues

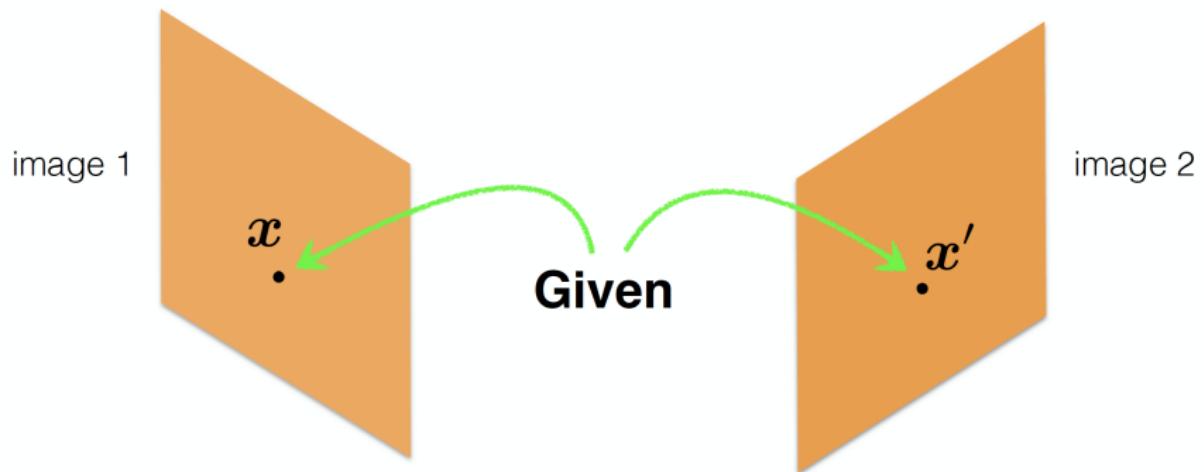
# Vers la reconstruction 3D



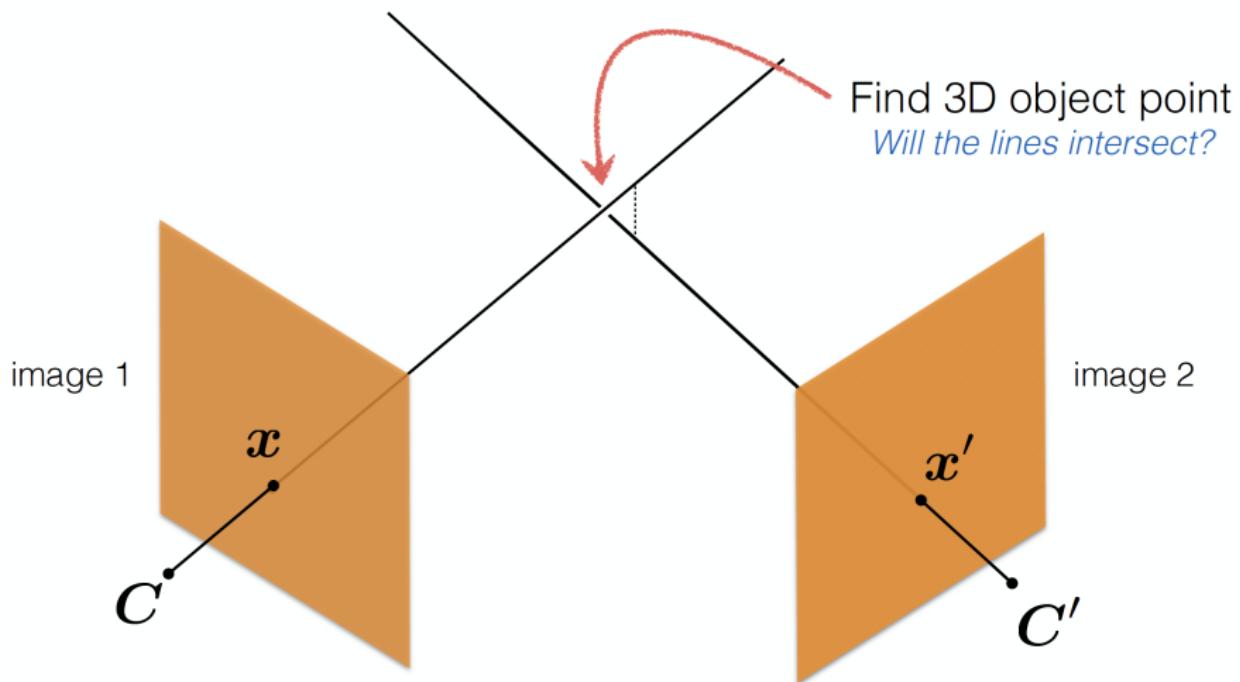
# Plan

- 1 Introduction
- 2 Projection perspective
- 3 Modèle de caméra
- 4 Retour sur la projection perspective
- 5 Calibration d'une caméra
- 6 Triangulation
- 7 Vision stéréoscopique
- 8 Structure from Motion
- 9 Conclusion

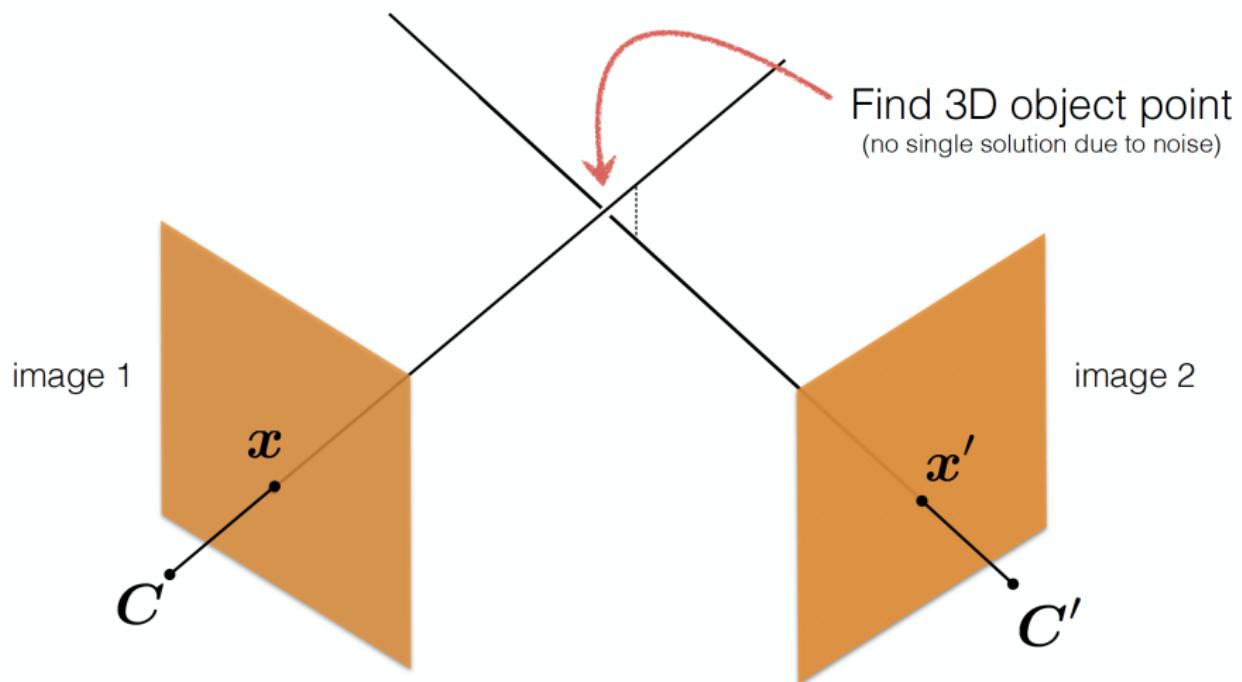
# Triangulation



# Triangulation



# Triangulation



# Triangulation

- Etant donné un ensemble de points correspondants (bruités) :

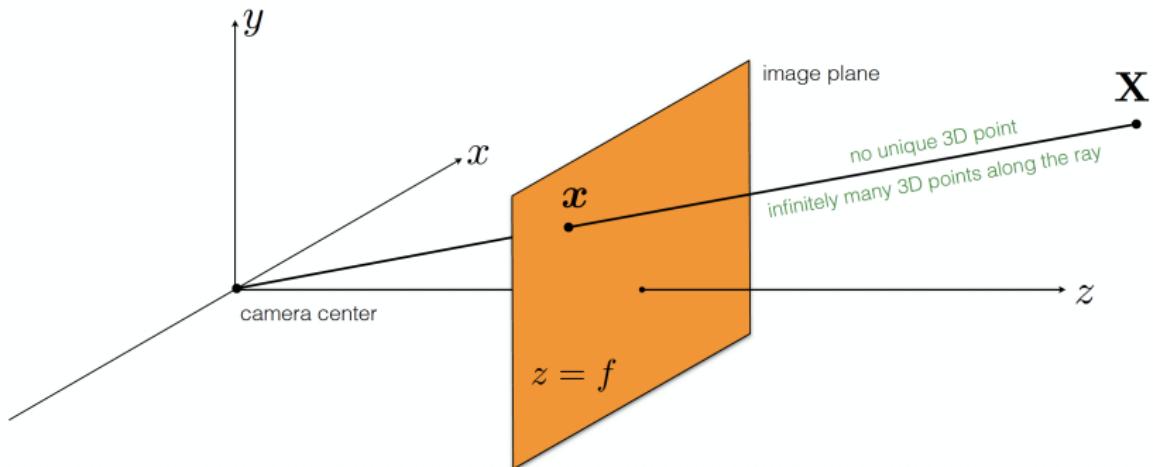
$$\{\mathbf{x}_i, \mathbf{x}'_i\}$$

- et les matrices de projection  $\mathbf{P}$  et  $\mathbf{P}'$
- Il faut estimer le point 3D  $\mathbf{X}$

$$\mathbf{x} = \mathbf{P}\mathbf{X}$$

Peut-on calculer  $\mathbf{X}$  à partir d'une seule correspondance ?

# Triangulation



# Triangulation

$$\mathbf{x} = \mathcal{P}\mathbf{X}$$

Peut-on calculer  $\mathbf{X}$  à partir de deux correspondances  $\mathbf{x}$  et  $\mathbf{x}'$  ?

- Oui si les mesures sont parfaites mais c'est rarement le cas.
- On va donc rechercher la meilleure correspondance.

# Triangulation

- $x = P\mathbf{X}$
- Même direction de rayon mais différence d'un facteur

$$x = \alpha P\mathbf{X}$$

- Produit vectoriel : le produit vectoriel de deux vecteurs de même direction est 0

$$\mathbf{x} \times P\mathbf{X} = 0$$

# Triangulation

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \alpha \begin{bmatrix} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \alpha \begin{bmatrix} \text{---} & \mathbf{p}_1^\top & \text{---} \\ \text{---} & \mathbf{p}_2^\top & \text{---} \\ \text{---} & \mathbf{p}_3^\top & \text{---} \end{bmatrix} \begin{bmatrix} | \\ \mathbf{X} \\ | \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \alpha \begin{bmatrix} \mathbf{p}_1^\top \mathbf{X} \\ \mathbf{p}_2^\top \mathbf{X} \\ \mathbf{p}_3^\top \mathbf{X} \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \times \begin{bmatrix} \mathbf{p}_1^\top \mathbf{X} \\ \mathbf{p}_2^\top \mathbf{X} \\ \mathbf{p}_3^\top \mathbf{X} \end{bmatrix} = \begin{bmatrix} y\mathbf{p}_3^\top \mathbf{X} - \mathbf{p}_2^\top \mathbf{X} \\ \mathbf{p}_1^\top \mathbf{X} - x\mathbf{p}_3^\top \mathbf{X} \\ x\mathbf{p}_2^\top \mathbf{X} - y\mathbf{p}_1^\top \mathbf{X} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$



# Triangulation

Comme  $\mathbf{x} \times \mathbf{P}\mathbf{X} = 0$

$$\mathbf{x} \times \mathbf{P}\mathbf{X} = \mathbf{0}$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \times \begin{bmatrix} \mathbf{p}_1^\top \mathbf{X} \\ \mathbf{p}_2^\top \mathbf{X} \\ \mathbf{p}_3^\top \mathbf{X} \end{bmatrix} = \begin{bmatrix} y\mathbf{p}_3^\top \mathbf{X} - \mathbf{p}_2^\top \mathbf{X} \\ \mathbf{p}_1^\top \mathbf{X} - x\mathbf{p}_3^\top \mathbf{X} \\ x\mathbf{p}_2^\top \mathbf{X} - y\mathbf{p}_1^\top \mathbf{X} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

La troisième ligne est une combinaison linéaire des deux précédentes.

# Triangulation

$$\begin{bmatrix} y\mathbf{p}_3^\top \mathbf{X} - \mathbf{p}_2^\top \mathbf{X} \\ \mathbf{p}_1^\top \mathbf{X} - x\mathbf{p}_3^\top \mathbf{X} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} y\mathbf{p}_3^\top - \mathbf{p}_2^\top \\ \mathbf{p}_1^\top - x\mathbf{p}_3^\top \end{bmatrix} \mathbf{X} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{A}_i \mathbf{X} = \mathbf{0}$$

On obtient alors un système d'équations linéaires.

# Triangulation

On concatène les points des deux images.

$$\begin{bmatrix} y\mathbf{p}_3^\top - \mathbf{p}_2^\top \\ \mathbf{p}_1^\top - x\mathbf{p}_3^\top \\ y'\mathbf{p}'_3^\top - \mathbf{p}'_2^\top \\ \mathbf{p}'_1^\top - x'\mathbf{p}'_3^\top \end{bmatrix} \mathbf{X} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

*sanity check! dimensions?*

$$\mathbf{A}\mathbf{X} = \mathbf{0}$$

Résolution avec une SVD (décomposition en valeurs singulières)

# Triangulation

Rappel : résolution aux moindres carrés.

$$\text{minimize } \|\mathbf{A}\mathbf{x}\|^2$$

$$\text{subject to } \|\mathbf{x}\|^2 = 1$$



$$\text{minimize } \frac{\|\mathbf{A}\mathbf{x}\|^2}{\|\mathbf{x}\|^2}$$

(Rayleigh quotient)

Solution is the eigenvector  
corresponding to smallest eigenvalue of

$$\mathbf{A}^\top \mathbf{A}$$

# Plan

- 1 Introduction
- 2 Projection perspective
- 3 Modèle de caméra
- 4 Retour sur la projection perspective
- 5 Calibration d'une caméra
- 6 Triangulation
- 7 Vision stéréoscopique
- 8 Structure from Motion
- 9 Conclusion

# Vision 3D

- Même en calibrant une caméra, nous ne connaissons pas la profondeur des points de l'image.
- L'information de profondeur manque pour avoir la 3D complète.
- A chaque point correspond une droite 3D dans la scène.
- Comment retrouver la 3ème dimension ?
  - ▶ En utilisant deux ou plusieurs vues, on peut évaluer la position 3D d'un objet.
  - ▶ Obtention d'une nouvelle vue par l'ajout d'une caméra ou en déplaçant la caméra.

# Reconstruction 3D

## Objectif

Reconstitution de la 3ème dimension qui a été perdue à cause de la projection sur le plan image.

Plusieurs stratégies :

- Mono-caméra : Shape from shading, from focus, Stéréo-photométrie
- Vision active : Lumière structurée, laser
- Multi-caméra et vision passive : stéréovision et reconstruction stéréoscopique, ...

# Reconstruction 3D

Mono-caméra : shape from focus

Principe : utilisation de l'effet de flou de défocalisation pour estimer la distance ou la forme des objets.

- On connaît les propriétés du système d'acquisition ; profondeur de champ, ouverture de l'objectif.
- Si la profondeur de champ est finie, le flou dépend de la distance entre l'objet et la caméra.



(a)



(b)



(c)

Source : Favaro et al

[http://www.cs.columbia.edu/CAVE/projects/shape\\_focus/](http://www.cs.columbia.edu/CAVE/projects/shape_focus/)

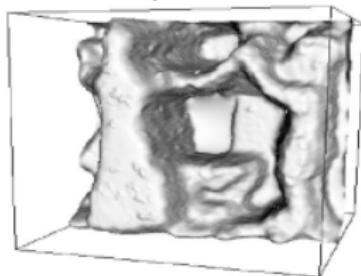
# Reconstruction 3D : shape from focus



mise au point distante



mise au point proche



reconstruction 3D



carte des distances

Source : Favaro et al

[http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL\\_COPIES/FAVARO1/dfdtutorial.html](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/FAVARO1/dfdtutorial.html)

# Reconstruction 3D

## Mono-caméra : shape from shading

Principe : utilisation des variations d'intensité dans l'image pour estimer la distance ou la forme des objets.

- On connaît les propriétés d'éclairage de la scène.
- Propriétés lambertiennes des surfaces : l'intensité de la lumière réfléchie dépend de son incidence.

# Reconstruction 3D

Mono-caméra : shape from shading

Input



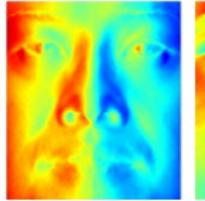
...



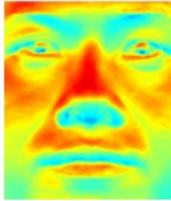
Estimated albedo



Estimated normals



X



y



Z

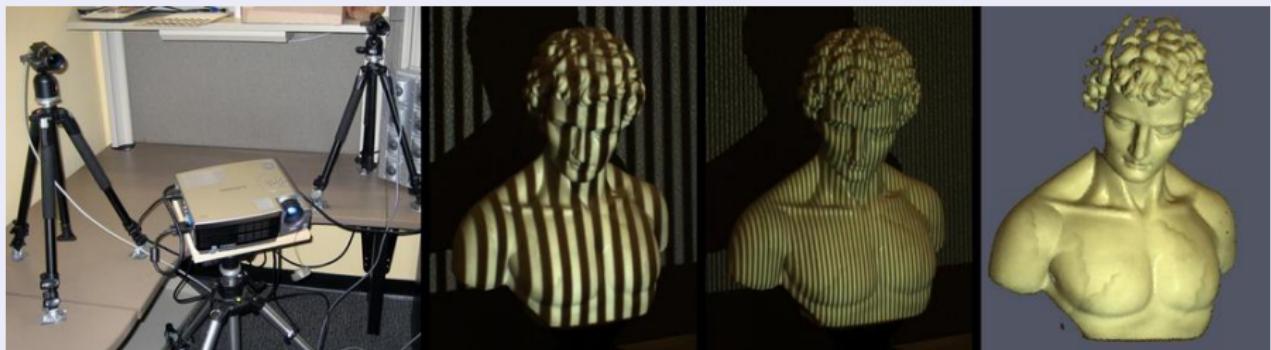
Integrated height map



# Reconstruction 3D

## Vision active : lumière structurée

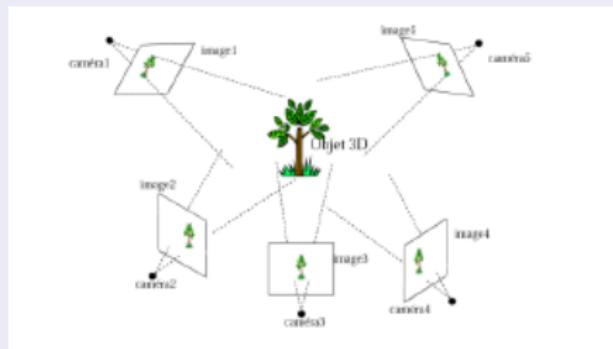
Contrôle de la structure de la source lumineuse pour faire apparaître le relief des objets.



# Reconstruction 3D : à partir de vues multiples

## Principe

Exploitation des différences existant entre les images acquises depuis différents points de vue.



# Vision stéréoscopique

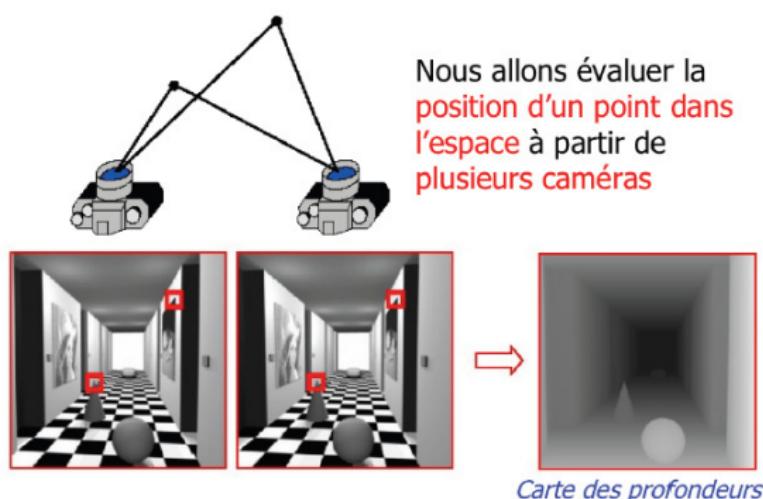


FIGURE – Source :A. Boucher

# Vision stéréoscopique

## Principe

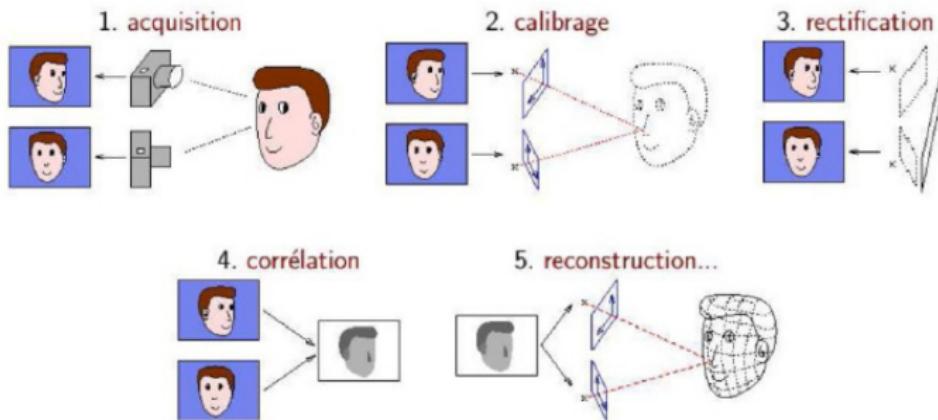
Reconstruction 3D de certains points de la scène à partir de deux images obtenues à partir de points de vue différents.

- Recherche de **points homologues** dans les deux images par **mise en correspondance**.
- Reconstruction 3D à partir des propriétés des points homologues.

# Vision stéréoscopique

- Calibration stéréoscopique
  - ▶ Géométrie épipolaire
  - ▶ Orientation relative des caméras
- Rectification d'images
  - ▶ Transformation d'une image pour simplifier la mise en correspondance.
- Mise en correspondance
  - ▶ Trouver les paires de points correspondants dans chaque image
- Reconstruction 3D
  - ▶ Triangulation pour reconstruire le modèle 3D

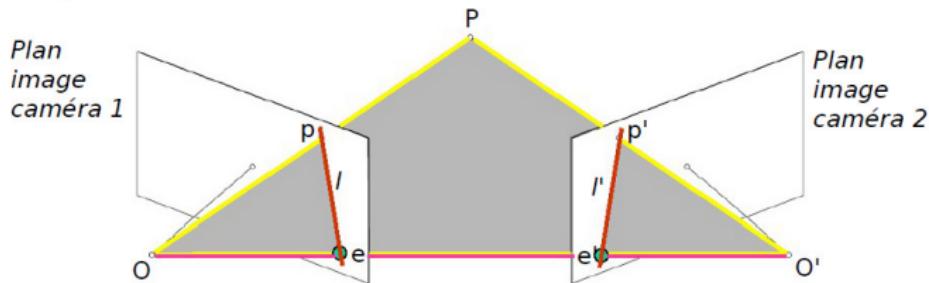
# Vision stéréoscopique



source : F. Devernay, Inria

# Vision stéréoscopique : Géométrie épipolaire

Modèle mathématique de géométrie décrivant les relations géométriques de différentes photos du même objet, prises de différents points d'observation.



## Géométrie épipolaire

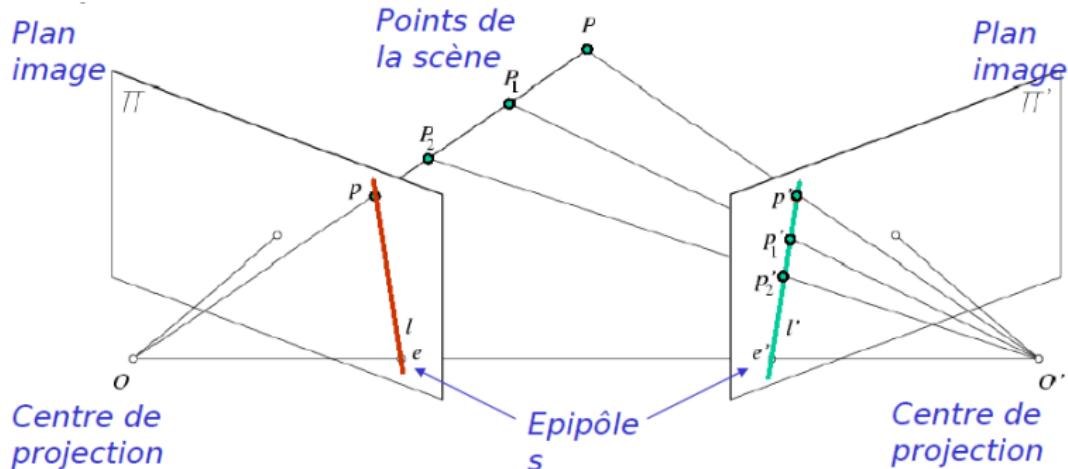
- 5 points dans le même plan :  $P, p, p', O$  et  $O'$  : plan épipolaire.
- $I$  et  $I'$  sont les droites épipolaires (conjuguées)
- Les points  $e$  et  $e'$  sont respectivement les épipôles gauche et droit : intersection de la droite reliant les 2 centres optiques avec les plans images.
- $e'$  est  $O$  vu dans l'image de droite.

# Vision stéréoscopique : Géométrie épipolaire

## Géométrie épipolaire

- Un **épipôle** est le centre de projection d'une caméra vue dans le plan image de l'autre caméra.
  - ▶ Une seule droite épipolaire passe par chaque point des images (sauf aux épipôles)
- Une **droite épipolaire** est formée par l'épipôle et un point dans l'image.
  - ▶ Les droites épipolaires passent toutes par les épipôles.

# Vision stéréoscopique : Géométrie épipolaire



Le point  $p'$ , correspondant de  $p$  dans l'image droite se situe sur la droite épipolaire  $e'p'$  (et réciproquement).

⇒ La géométrie épipolaire induit une contrainte forte sur le problème de la mise en correspondance.

# Vision stéréoscopique : Géométrie épipolaire

- Le correspondant d'un point de l'image gauche dans l'image droite est constraint par la droite épipolaire.
  - ▶ Il s'agit de la **contrainte épipolaire**.
  - ▶ Hypothèse d'aucune autre distorsion que la projection perspective.
  - ▶ Cela limite l'espace de recherche pour trouver le correspondant d'un point.
- Lorsque les deux plans des caméras sont parallèles (**configuration canonique**) :
  - ▶ Les droites épipolaires sont parallèles.
  - ▶ Les épipoles sont à l'infini.

# Vision stéréoscopique : Géométrie épipolaire

Exemples de droites épipolaires



# Vision stéréoscopique : Géométrie épipolaire

## Paramètres de la géométrie

- Problème : connaître les épipôles et les paramètres de la géométrie épipolaire.
- Solution :
  - ▶ Calibration d'un système à deux caméras fixes.
    - ★ Calibrer chaque caméra.
    - ★ Connaître la transformation qui permet de passer d'une caméra vers l'autre caméra : **calibration stéréoscopique** (caractérisation de la géométrie relative entre deux caméras)
    - ★ Calcul des droites épipolaires.
  - ▶ Eventuellement, étape de rectification.
  - ▶ Etape de mise en correspondance.

# Vision stéréoscopique : calibration

## Problème

Calculer les paramètres d'une caméra en utilisant uniquement des points mis en correspondance entre deux images d'une même scène, prise sous des points de vue différents.

Exprime la relation entre les paramètres externes des deux caméras, pas entre les caméras et la scène.

## 2 cas : Paramètres internes

- inconnus  $\Rightarrow$  matrice fondamentale (dans la repère image).
- connus  $\Rightarrow$  matrice essentielle (dans le repère caméra).

La géométrie épipolaire peut être entièrement définie par ces 2 matrices

The Fundamental Matrix Song :

<https://www.youtube.com/watch?v=DgGV3l82NTk>

# Modélisation de la géométrie épipolaire

- Un point  $M$  de la scène peut être exprimé dans les repères des 2 caméras :  $(X, Y, Z, 1)^T$  et  $(X', Y', Z', 1)^T$
- Les deux repères sont liés par une rotation et une translation 3D.

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = \mathbf{A} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \text{ avec } \mathbf{A} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- On peut exprimer la position du point dans le repère droit (D) en utilisant les coordonnées gauches.

$$\begin{cases} X' = r_{11}X + r_{12}Y + r_{13}Z + t_x \\ Y' = r_{21}X + r_{22}Y + r_{23}Z + t_y \\ Z' = r_{31}X + r_{32}Y + r_{33}Z + t_z \end{cases}$$

# Modélisation de la géométrie épipolaire

- La projection de  $M$  dans la caméra droite génère le point  $m' = (x', y')$  ( $f_d = 1$ ) :

$$\begin{cases} x' = \frac{X'}{Z'} = \frac{r_{11}X + r_{12}Y + r_{13}Z + t_x}{r_{31}X + r_{32}Y + r_{33}Z + t_z} \\ y' = \frac{Y'}{Z'} = \frac{r_{21}X + r_{22}Y + r_{23}Z + t_y}{r_{31}X + r_{32}Y + r_{33}Z + t_z} \end{cases}$$

- En utilisant la relation  $X = x.Z$  et  $Y = y.Z$  (focale  $f_g = 1$ ) on arrive à :

$$\begin{cases} x' = \frac{Z(r_{11}x + r_{12}y + r_{13}) + t_x}{Z(r_{31}x + r_{32}y + r_{33}) + t_z} \\ y' = \frac{Z(r_{21}x + r_{22}y + r_{23}) + t_y}{Z(r_{31}x + r_{32}y + r_{33}) + t_z} \end{cases}$$

- Si on note  $\mathbf{r}_i$  la ième ligne de la matrice de rotation  $\mathbf{R}$  et  $\mathbf{m} = (x, y, 1)^T$  :

$$\begin{cases} x' = \frac{Z\mathbf{r}_1\mathbf{m} + t_x}{Z\mathbf{r}_3\mathbf{m} + t_z} \\ y' = \frac{Z\mathbf{r}_2\mathbf{m} + t_y}{Z\mathbf{r}_3\mathbf{m} + t_z} \end{cases}$$

# Relations entre les projections

## Interprétation

La projection d'un point 3D dans l'image **droite** peut s'exprimer en fonction de :

- De la projection de ce point dans l'image **gauche**
- De la rotation et de la translation entre les 2 caméras.
- De la profondeur Z (projection perspective)
- En formalisant encore plus, à partir des expressions précédentes, on a :

$$\begin{cases} Z = \frac{t_x - x't_z}{x'\mathbf{r}_3\mathbf{m} - \mathbf{r}_1\mathbf{m}} \\ Z = \frac{t_y - y't_z}{y'\mathbf{r}_3\mathbf{m} - \mathbf{r}_2\mathbf{m}} \end{cases}$$

# Relations entre les projections

- On a donc :

$$\frac{t_x - x't_z}{x'\mathbf{r}_3\mathbf{m} - \mathbf{r}_1\mathbf{m}} = \frac{t_y - y't_z}{y'\mathbf{r}_3\mathbf{m} - \mathbf{r}_2\mathbf{m}}$$

- Après développement :

$$x'(t_z\mathbf{r}_2 - t_y\mathbf{r}_3)\mathbf{m} - y'(t_z\mathbf{r}_1 - t_x\mathbf{r}_3)\mathbf{m} - (t_x\mathbf{r}_2 - t_y\mathbf{r}_1)\mathbf{m} = 0$$

- Ceci correspond à l'équation d'une droite **I'** dans l'image droite et peut être écrit  $x'a + y'b + c = 0$  avec :

$$\begin{cases} a = (t_z\mathbf{r}_2 - t_y\mathbf{r}_3)\mathbf{m} \\ -b = (t_z\mathbf{r}_1 - t_x\mathbf{r}_3)\mathbf{m} \\ -c = (t_x\mathbf{r}_2 - t_y\mathbf{r}_1)\mathbf{m} \end{cases}$$

# Relation point → droite épipolaire

## relation point-droite

- A un point dans une image correspond une droite dans l'autre image.
- Cette droite est la **droite épipolaire** et ne dépend que de la relation inter-caméras.
- Relation non réversible
- Les coefficients de cette droite s'écrivent :

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

# Matrice Essentielle E

- Cette relation est formalisée par la **matrice Essentielle E**, qui à un point  $\mathbf{p} = (x, y, 1)^T$  associe une droite  $\mathbf{l}' = (a', b', c')^T$  dans le repère caméra :

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{E} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

- E** s'écrit :

$$\mathbf{E} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

# Matrice Fondamentale F

- La relation épipolaire peut aussi être exprimée en coordonnées-image.
- On utilise alors la **matrice fondamentale (F)** qui associe un point à sa droite épipolaire :

$$l' = Fm$$

$$l = F^T m'$$

- $F$  et  $E$  sont liées par les paramètres intrinsèques (matrice  $K$ ) :

$$E = K'^T FK$$

$$F = K'^{-T} E K^{-1}$$

- $F$  peut être déterminée uniquement à partir de correspondances entre points-images, sans connaissance des caméras.

# Contrainte épipolaire

- Tous les points stéréo-correspondants respectent cette contrainte.
- Elle peut s'exprimer de 2 façons selon que l'on se place dans le repère caméra ou le repère image :
  - ▶ En coordonnées image :  $\mathbf{m}'^T \cdot \mathbf{F} \cdot \mathbf{m} = 0$
  - ▶ En coordonnées caméra :  $\mathbf{p}'^T \cdot \mathbf{E} \cdot \mathbf{p} = 0$
- **Interprétation : tout point droit correspondant à un point gauche doit se trouver sur la ligne épipolaire correspondante.**

# Vision stéréoscopique : calibration

## Cas des caméras calibrées

- Les matrices de calibration de chaque caméra sont connues :

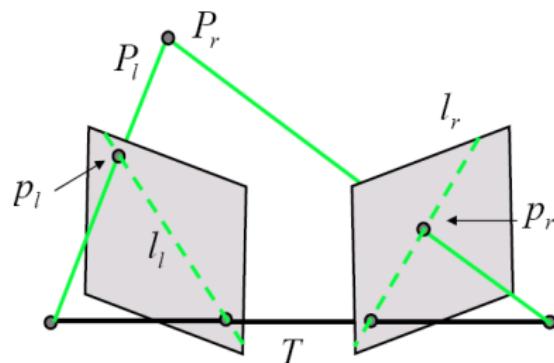
$$P_i = MP_e = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \cdot P_e$$

- On peut donc calculer la matrice essentielle (E) pour ces deux caméras :

- ▶ Relation entre les points de l'image définies en coordonnées du référentiel caméra et les droites épipolaires.

$\hat{p}_l^T E \hat{p}_r = 0$  avec  $\hat{p}_l$  et  $\hat{p}_r$  les coordonnées des pixels dans le référentiel caméra.

# Vision stéréoscopique : calibration



## Cas des caméras calibrées

- Matrice essentielle :  $\hat{p}_l^T E \hat{p}_r = 0$
- Les droites épipolaires sont données par :
  - ▶  $l_l = E \hat{p}_r$
  - ▶  $l_r = E^T \hat{p}_l$

# Vision stéréoscopique : calibration

Cas des caméras calibrées : mire de calibration

On connaît pour chaque point :

- sa position  $(x_1, y_1)$  en pixels dans l'image gauche.
- sa position  $(x_2, y_2)$  en pixels dans l'image droite.
- sa position réelle  $(x, y, z)$  en 3D dans la scène.

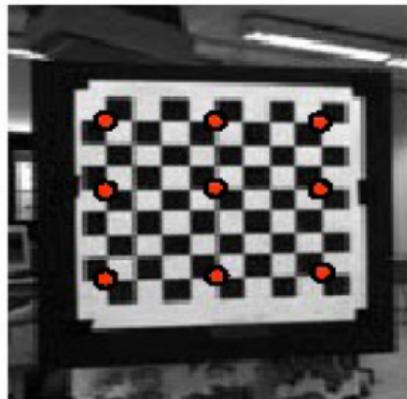


Image caméra gauche

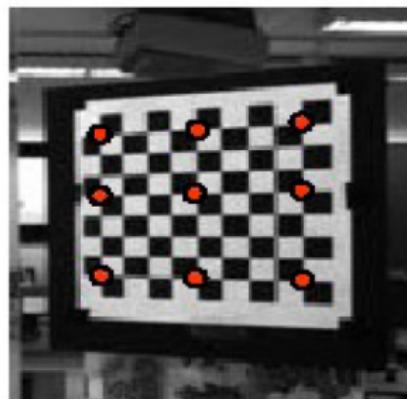


Image caméra droite

# Vision stéréoscopique :calibration

## Cas des caméras non calibrées

- Si les matrices de calibration de chaque caméra ne sont pas connues, la relation entre les coordonnées dans le référentiel image et celles dans le référentiel caméra n'est pas connue.
- Calcul de la **matrice fondamentale** pour ces deux caméras.
  - ▶ Relation entre les points de l'image définies en coordonnées du référentiel image et les droites épipolaires :

$$\overline{p_l}^T F \overline{p_r} = 0$$

# Vision stéréoscopique : calibration

## Cas des caméras non calibrées : matrice fondamentale

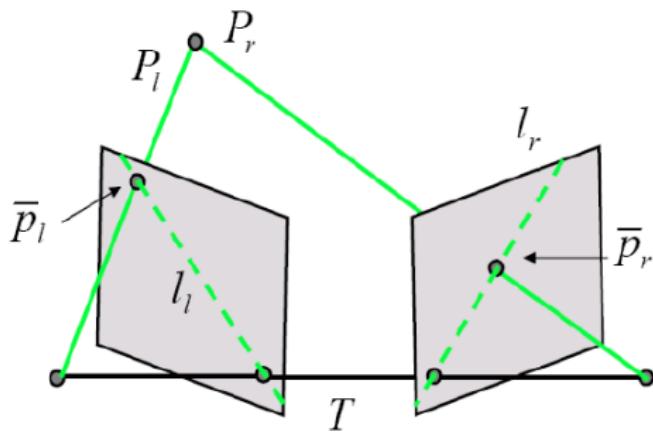
- $\hat{p}_l^T E \hat{p}_r = 0$ , avec  $\hat{p}_l$  et  $\hat{p}_r$  les coordonnées des pixels dans le référentiel caméra.
- $\overline{p_l} = K_l \hat{p}_l$  et  $\overline{p_r} = K_r \hat{p}_r$ , avec  $K$ , matrice de calibration de la camera (camera  $\rightarrow$  image)
- $\overline{p_l}^T K_l^T E K_r \overline{p_r} = 0$
- $\overline{p_l}^T F \overline{p_r} = 0$ , avec  $\overline{p_l}$  et  $\overline{p_r}$ , les coordonnées des pixels dans le référentiel image.

# Vision stéréoscopique : calibration

Cas des caméras non calibrées : matrice fondamentale

Les droites épipolaires sont données par :

- $l_l = F \bar{p}_r$
- $l_r = F^T \bar{p}_l$



# Vision stéréoscopique : calibration

Cas des caméras non calibrées : mire de calibration

On connaît pour chaque point :

- sa position  $(x_1, y_1)$  en pixels dans l'image gauche.
- sa position  $(x_2, y_2)$  en pixels dans l'image droite.
- On ne connaît pas sa position réelle.

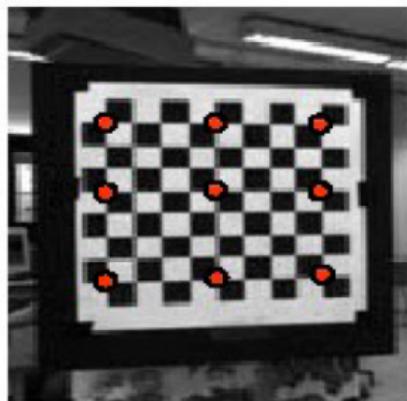


Image caméra gauche

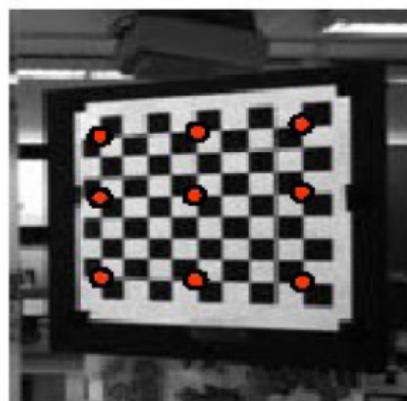


Image caméra droite

# Vision stéréoscopique : calibration

## Cas des caméras non calibrées : estimation de la matrice fondamentale

- Procédure de calibration faible :
  - ▶ Calibration des 2 caméras sans connaître et sans rechercher les paramètres extrinsèques et intrinsèques de chacune des caméras.
  - ▶ Estimer  $F$  sans connaître les deux  $K$  (ou  $M$ ).
- $F$  : matrice  $3 \times 3$  de rang 2 contenant 8 inconnues.
- Résolution avec plusieurs correspondances :
$$\min_F \sum (\overline{p_l}^T F \overline{p_r})^2 \text{ avec la contrainte } \det(F) = 0$$
- Algorithme de décomposition en valeurs singulières (SVD)

# Vision stéréoscopique : calibration

## Cas des caméras non calibrées : estimation de la matrice fondamentale

Méthode de base : Algorithme des 8 points.

- Algorithme simple et populaire
- Formulation linéaire.
- Soient  $N$  points (avec  $N \leq 8$ ) mis en correspondance. On a l'équation :

$$p^T \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} q = 0$$

- Une correspondance = une contrainte.

# Vision stéréoscopique : calibration

## Estimation de la matrice fondamentale - Méthode de base

$$A \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_9 \end{pmatrix} = 0$$

avec :

$$A = \begin{bmatrix} p_1 q_1 & p_1 q_2 & p_1 q_3 & p_2 q_1 & p_2 q_2 & p_2 q_3 & p_3 q_1 & p_3 q_2 & p_3 q_3 \\ \vdots & \vdots \\ \vdots & \vdots \end{bmatrix}$$

Résolution aux moindres carrés : déterminer  $\hat{f}$  tel que :

$$\hat{f} = \operatorname{argmin}_f \|Af\|^2 \text{ subject to } \|f\| = 1$$

# Vision stéréoscopique : calibration

## Estimation de la matrice fondamentale - Méthode de base

Mais la méthode précédente néglige une contrainte de la matrice fondamentale qui doit être singulière (toutes les droites épipolaires passent par les éipoles,  $e'^T F = 0$ ).

⇒

Etant donné une matrice  $F$  (non singulière, par exemple celle trouvée par la méthode précédente), trouver  $\hat{f}'$  qui minimise  $\|f - f'\|$ .

- Résolution de ce problème par une décomposition en valeurs singulières.

# De la calibration à la reconstruction

La reconstruction stéréo est le procédé qui permet, en déterminant des paires de points homologues dans les deux images, d'estimer par triangulation la troisième dimension.

Plusieurs étapes :

- Identification des points d'intérêts (primitives)
- Recherche de correspondances entre les primitives.
- Reconstruction 3D par triangulation

Deux points  $m$  et  $m'$  sont dits homologues s'ils correspondent aux deux projections d'un même point  $M$  sur les deux images.

# De la calibration à la reconstruction

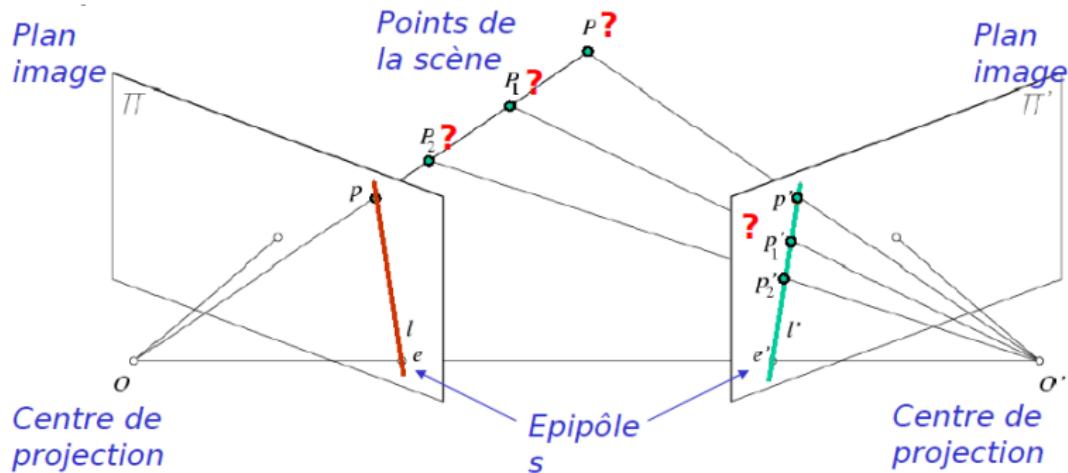
## Deux méthodes de reconstruction

- **Eparse** : seuls quelques points sont utilisés pour la mise en correspondance.
- **Dense** : recherche des homologies de tous les points des images et identification des zones occultées.

# Vision stéréoscopique : mise en correspondance

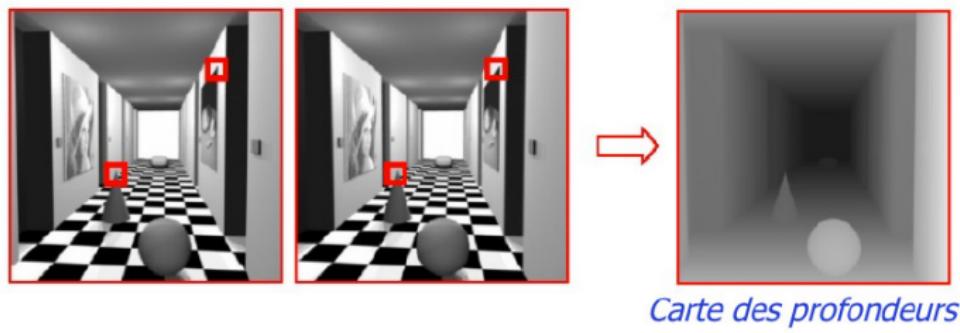
## Problème

A partir d'un point dans l'image gauche et de la géométrie épipolaire, quel est le bon point correspondant dans l'image droite ?



# Vision stéréoscopique : mise en correspondance

- Si on retrouve l'ensemble des correspondances, on peut calculer une carte de profondeur de la scène.
- Carte de disparités entre les correspondances (distance  $(p1, p2)$ ).

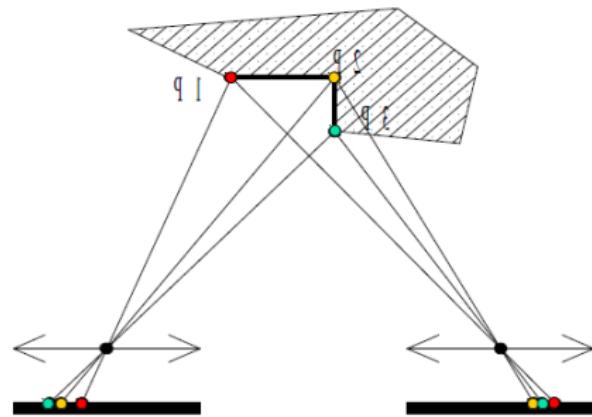


Carte des profondeurs

# Vision stéréoscopique : mise en correspondance

## Difficultés

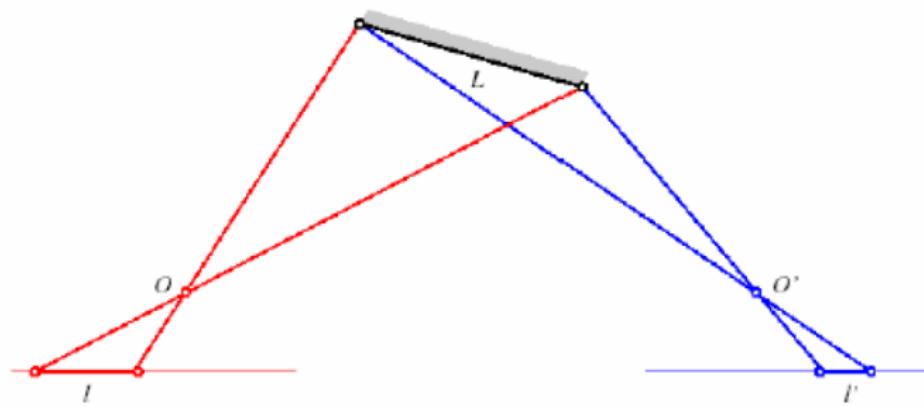
Les points ne sont pas forcément dans le même ordre dans les deux images.



# Vision stéréoscopique : mise en correspondance

## Difficultés

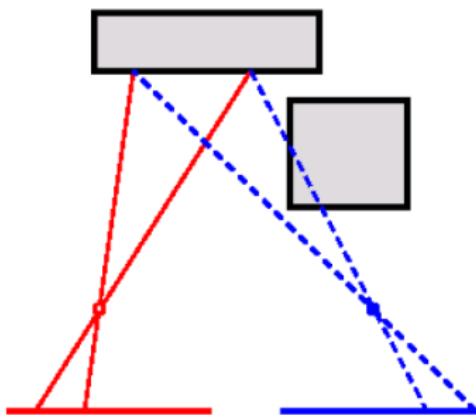
Les tailles et les distances ne sont pas les mêmes d'une image à l'autre



# Vision stéréoscopique : mise en correspondance

## Difficultés

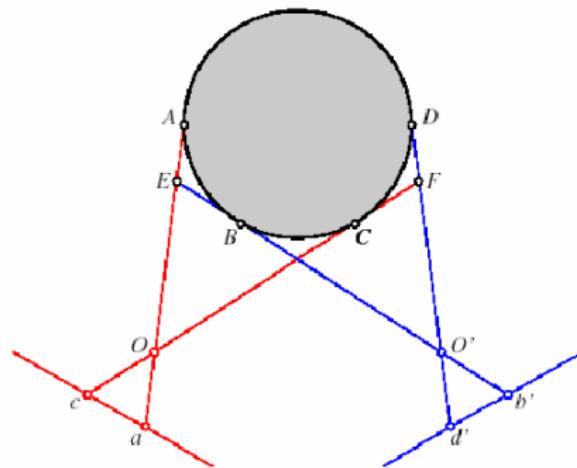
Occultations : des objets ou parties d'objets peuvent être cachés. La correspondance n'existe donc pas dans ce cas.



# Vision stéréoscopique : mise en correspondance

## Difficultés

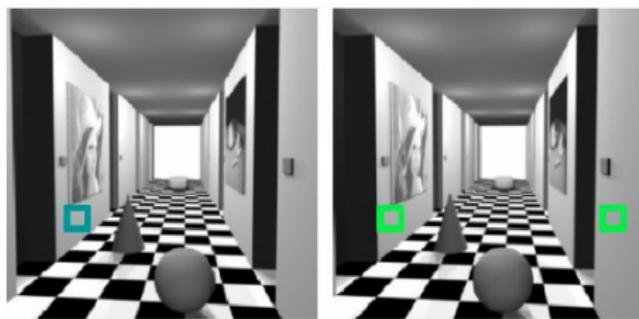
Dans le cas d'objets sphériques ou curvilignes, les arêtes percues ne correspondent pas d'une image à l'autre.



# Vision stéréoscopique : mise en correspondance

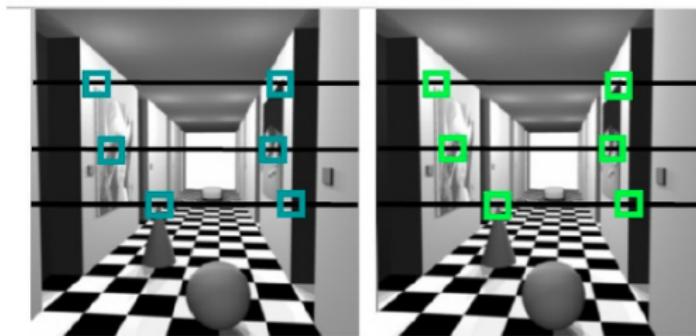
## Difficultés

Correspondance difficile voire impossible dans les zones faiblement texturées.



# Vision stéréoscopique : mise en correspondance

- Si les images sont rectifiées, alors la recherche des correspondances est beaucoup plus facile.
- Recherche des zones similaires sur la même ligne.
- La distance ( $x_2 - x_1$ ) entre les deux correspondances est la disparité : indication de la profondeur du point dans la scène.



# Vision stéréoscopique : mise en correspondance

## Rectification

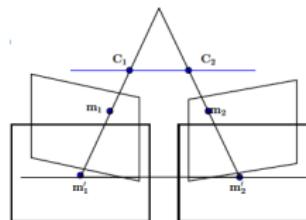
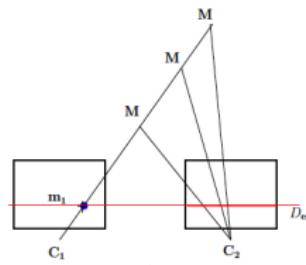
**Si :**

- Epipolaires parallèles entre elles.
- Epipolaires parallèles à l'axe des  $x$ .
- Choix des repères images tel que  $(x_1, y_1)$  dans  $\pi_1$  ait pour épipolaire conjuguée  $y_2 = y_1$  dans  $\pi_2$ , i.e.  $e_1$  et  $e_2$  sont à l'infini

**Alors**

La mise en correspondance se ramène à la recherche du correspondant sur la même ligne de base.

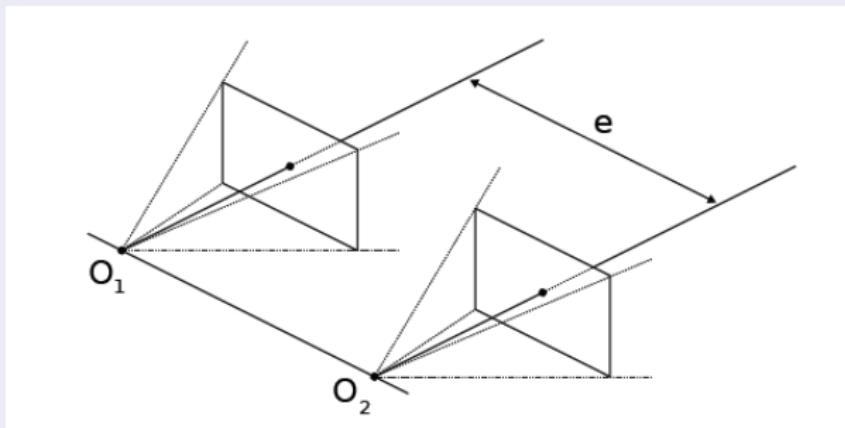
La transformation des images pour atteindre cette configuration s'appelle la **rectification**.



# Vision stéréoscopique : mise en correspondance

Rectification : configuration canonique d'une paire de stéréovision

Configuration qui simplifie la mise en correspondance



Caméras identiques, axes optiques parallèles.

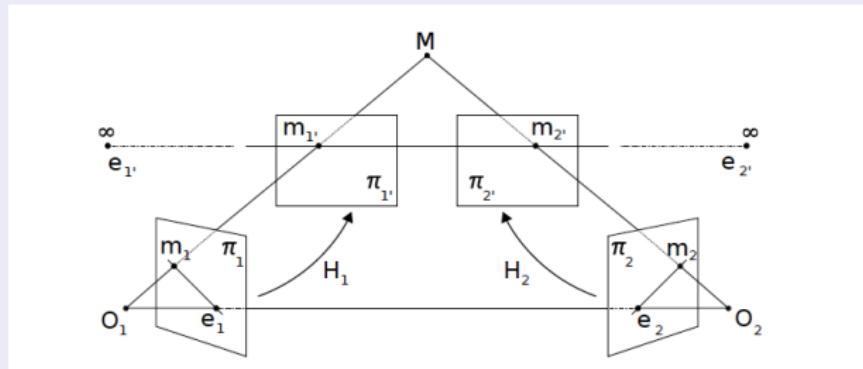
Epipoles à l'infini, droites épipolaires associées à deux points homologues sont confondues. On notera  $\bar{F}$  la matrice fondamentale de la configuration canonique.

# Vision stéréoscopique : mise en correspondance

## Rectification : objectif

Revenir à la configuration canonique :

- Transformer les images acquises par une paire de stéréovision non canonique par des homographies.
- Estimer des homographies qui permettent de conserver la plus grande partie commune sur les images.

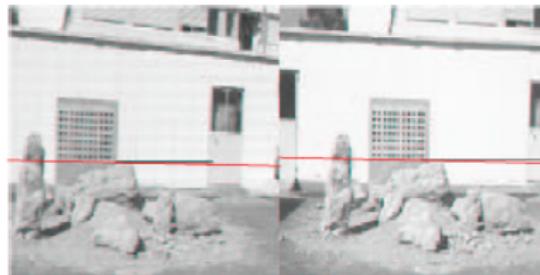


On peut montrer que les deux homographies  $\mathcal{H}_1$  et  $\mathcal{H}_2$  doivent vérifier  $F = \mathcal{H}_2^T \tilde{F} \mathcal{H}_1$ .

# Vision stéréoscopique : mise en correspondance

## Rectification

Avant rectification

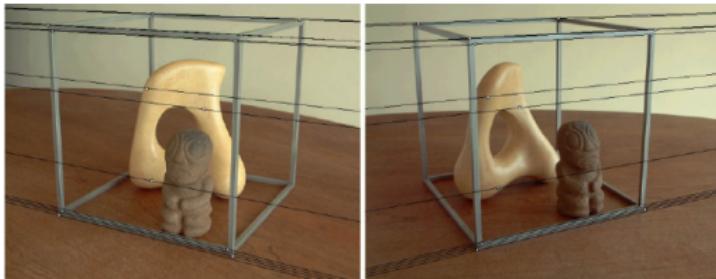


Après rectification

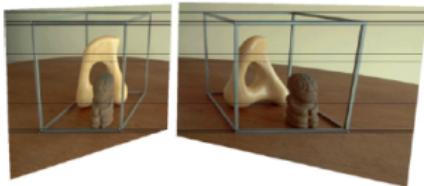


# Vision stéréoscopique : mise en correspondance

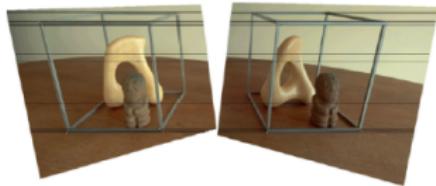
## Rectification évitant les déformations



Paire d'images acquise depuis deux points de vue très différents



avec déformation



déformation minimale

Source : Loop et Zhang

<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.5603>

# Vision stéréoscopique : mise en correspondance

## Hypothèses

- Les points de la scène sont visibles, en général, des deux points de vue.
- Les points correspondants sont similaires dans les images.

## Deux approches

- Correspondance de caractéristiques :
  - ▶ Ensemble éparse de correspondance
  - ▶ Points, coins, ...
- Corrélation de voisinage.
  - ▶ Permet un ensemble dense de correspondance.

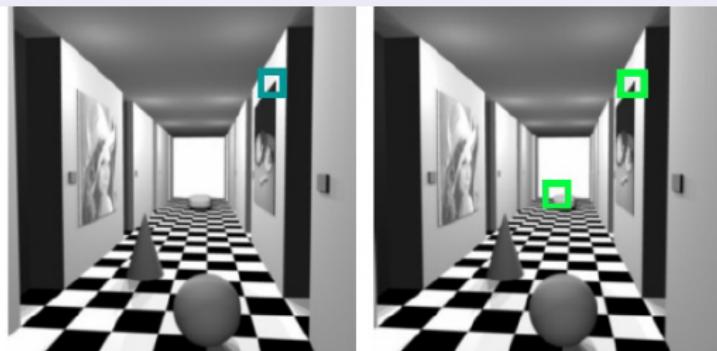
# Vision stéréoscopique : mise en correspondance

## Approche basée sur la corrélation

- On sélectionne une fenêtre définissant un voisinage dans l'image gauche.
- On cherche la fenêtre de voisinage correspondante dans l'image de droite.
  - ▶ Si on connaît la géométrie épipolaire, recherche le long de la droite épipolaire, sinon dans toute l'image.
  - ▶ Concentration de recherche dans les environs du point observé.

# Vision stéréoscopique : mise en correspondance

Approche basée sur la corrélation



$$\sum ( \boxed{\text{[Patch]}} - \boxed{\text{[Patch]}} )^2 = ssd$$

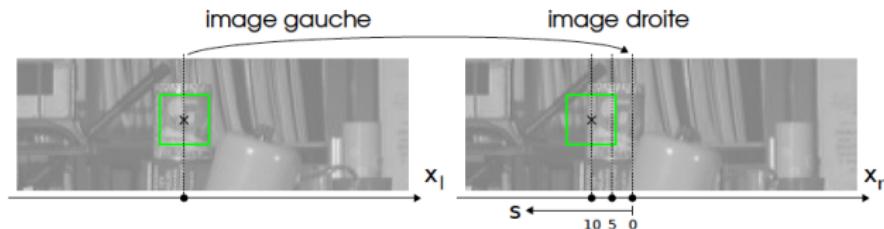
$$\sum ( \boxed{\text{[Patch]}} - \boxed{\text{[Patch]}} )^2 = ssd$$

# Vision stéréoscopique : mise en correspondance

## Bilan : contraintes pour la recherche des points homologues

- Contrainte épipolaire et rectification.
- Contrainte d'unicité.  
A chaque pixel de l'image de référence est associé au maximum un pixel de l'autre image, et inversement.
- Contrainte d'ordre (pas forcément vérifiée).
- Contraintes liées à la photométrie : similarité des niveaux de gris dans les voisinages (non vérifiée à proximité des discontinuités)

# Vision stéréoscopique : mise en correspondance



- Définition d'un voisinage  $\mathcal{V}(x, y)$  centré sur chaque pixel.
- Un voisinage dans l'image de droite est repéré par son décalage  $s$ .
- Définition d'une fonction de similarité ou de dissimilarité entre deux voisinages, l'un à gauche et l'autre à droite.
- La fonction a pour paramètres :
  - ▶ l'abscisse  $x_l$  du centre du voisinage gauche.
  - ▶ l'ordonnée commune  $y$  des centres des voisinages.
  - ▶ le décalage  $s$  entre les abscisses des voisinages.
- Un extremum indique que les voisinages sont similaires.

# Vision stéréoscopique : mise en correspondance

## Les différentes fonctions de dissimilarité

- SAD : somme des différences absolues

$$SAD(x_l, y, s) = \sum_{i=-w_x}^{w_x} \sum_{j=-w_y}^{w_y} |I_l(x_l + i, y + j) - I_r(x_l + i - s, y + j)|$$

- SSD : somme des différences au carré

$$SSD(x_l, y, s) = \sum_{i=-w_x}^{w_x} \sum_{j=-w_y}^{w_y} (I_l(x_l + i, y + j) - I_r(x_l + i - s, y + j))^2$$

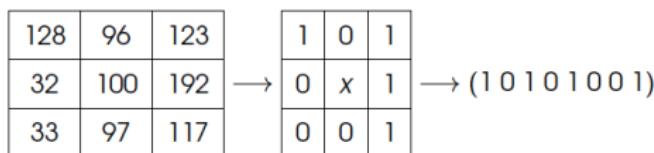
- NNC : Corrélation croisée normalisée

$$NNC(x_l, y, s) = \frac{\sum_{i=-w_x}^{w_x} \sum_{j=-w_y}^{w_y} (I_l(x_l + i, y + j) - \mu_{I_l}^{x_l})(I_r(x_l + i - s, y + j) - \mu_{I_r}^{x_r})}{\sqrt{\sum_{i=-w_x}^{w_x} \sum_{j=-w_y}^{w_y} (I_l(x_l + i, y + j))^2} \cdot \sqrt{\sum_{i=-w_x}^{w_x} \sum_{j=-w_y}^{w_y} (I_r(x_l + i - s, y + j))^2}}$$

# Vision stéréoscopique : mise en correspondance

## Les différentes fonctions de dissimilarité

- Transformée census : association d'une chaîne binaire à un voisinage
  - ▶ Un bit de la chaîne vaut 1 si le niveau de gris du pixel correspondant est supérieur ou égal à celui du pixel central et 0 sinon.



128	96	123
32	100	192
33	97	117

→

1	0	1
0	x	1
0	0	1

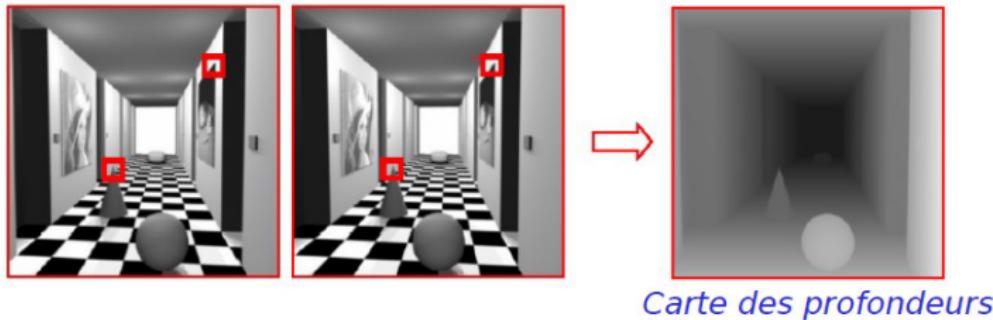
→ (10101001)

On prend ensuite la distance de Hamming qui compte le nombre de bits différents dans les deux chaînes.

# Vision stéréoscopique : mise en correspondance

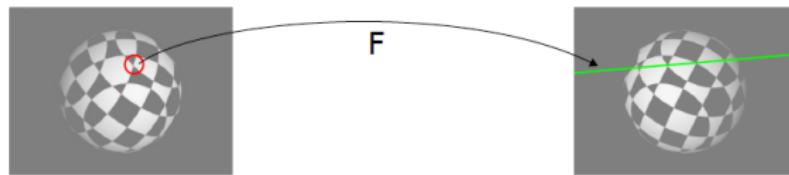
## Carte de profondeur

- Avec l'ensemble des correspondances, nous pouvons calculer une carte de profondeur de la scène.
- En fait, carte de disparité entre les correspondances.

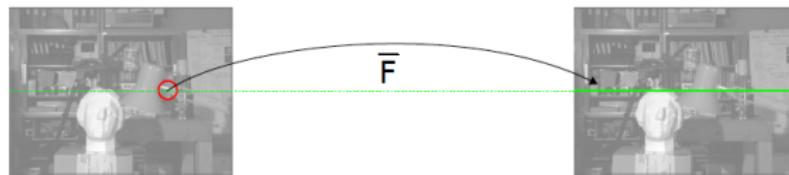


# Vision stéréoscopique : mise en correspondance

Le point homologue  $m_2$  dans l'image droite d'un point  $m_1$  dans l'image de gauche est situé sur la droite épipolaire de ce dernier.



configuration quelconque



configuration canonique

# Vision stéréoscopique : mise en correspondance

## Disparité

La disparité  $\delta$  est la distance entre les deux projetées du point  $M$  dans  $\pi_1$  et  $\pi_2$  :

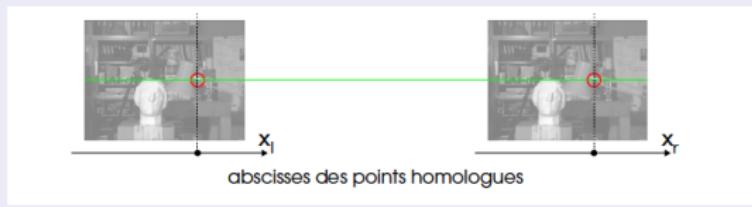
$$\delta = d(m_1, m_2)$$

## Cas des images rectifiées

Quand les images sont rectifiées ( $y_1 = y_2$ ), la disparité est donné par :

$$\delta = x_2 - x_1$$

, différence entre les abscisses de deux points homologues.



# Vision stéréoscopique : mise en correspondance

## Relation Disparité - Profondeur

Rappel : nous avons vu que nous avions :

$$Z_1 = \frac{t_x - t_z x_2}{\mathbf{r}_3 \cdot \mathbf{p} x_2 - \mathbf{r}_1 \cdot \mathbf{p}}$$

Si les images sont rectifiées on a comme matrice de changement de repère :

$$M = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

On a déduit :

$$Z_1 = \frac{t_x}{x_2 - \mathbf{r}_1 \cdot \mathbf{p}} = \frac{t_x}{x_2 - x_1} = \frac{t_x}{\delta}$$

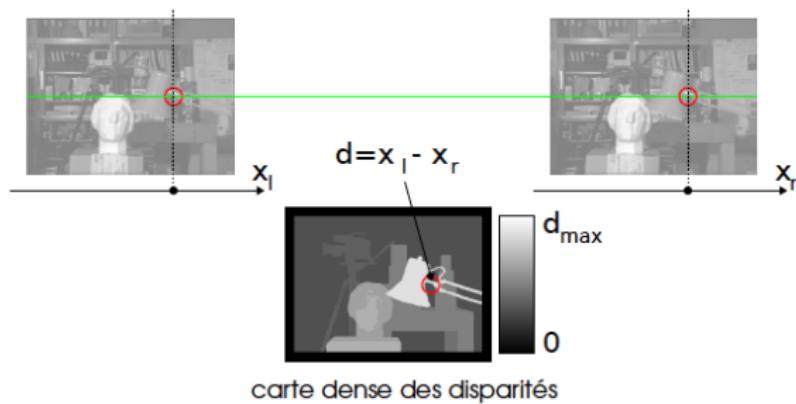
La profondeur est inversement proportionnelle à la disparité.

# Vision stéréoscopique : mise en correspondance

## Carte des disparités

Représentation de la disparité sous la forme d'une image :

- Carte éparsé : une disparité est associée à quelques pixels.
- Carte dense : une disparité est associée à chaque pixel.



# Vision stéréoscopique :deux types de production

## Stéréovision dense

Production d'une carte de profondeur donnant  $Z$  pour tout point de l'image.

- Coût calculatrice élevé
- Nécessité d'un post-traitement pour l'analyse de scène.

## Stéréovision éparsé

La profondeur est calculée uniquement sur les points caractéristiques.

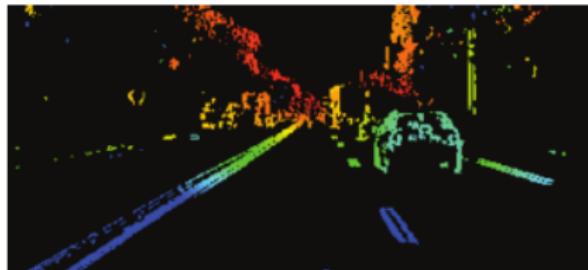
- Volume de données à traiter réduit.
- Performances liées à celle du détecteur.

# Vision stéréoscopique :deux types de production

- ▶ Stéréovision dense



- ▶ Stéréovision éparse



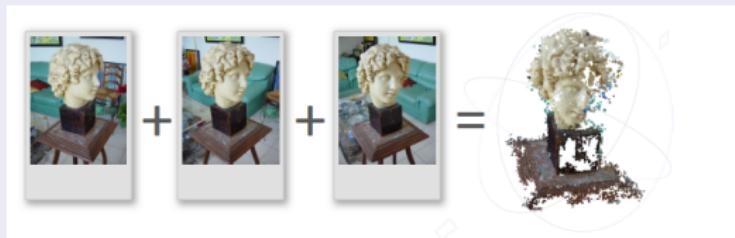
# Plan

- 1 Introduction
- 2 Projection perspective
- 3 Modèle de caméra
- 4 Retour sur la projection perspective
- 5 Calibration d'une caméra
- 6 Triangulation
- 7 Vision stéréoscopique
- 8 Structure from Motion
- 9 Conclusion

# Reconstruction 3D

Slides from Princeton : SFMedu : Multi-view 3D Reconstruction for Dummies. <http://vision.princeton.edu/courses/SFMedu/>

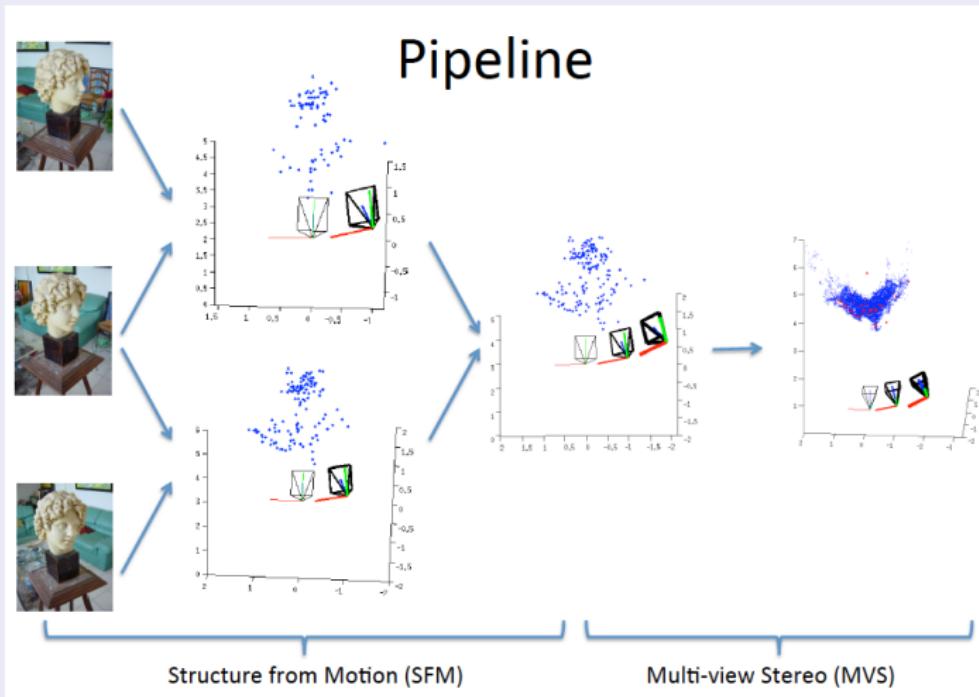
## Structure from motion



- Structure : nuage de points 3D représentant la scène.
- Mouvement : position de la caméra et son orientation.
- SFM : construire le nuage de points à partir de caméras en déplacement (plusieurs vues).

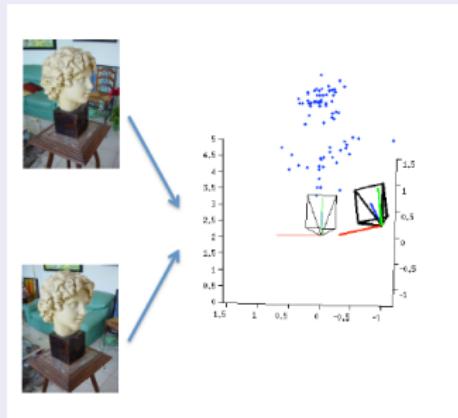
# Reconstruction 3D

## Structure from motion : Principe



# Reconstruction 3D

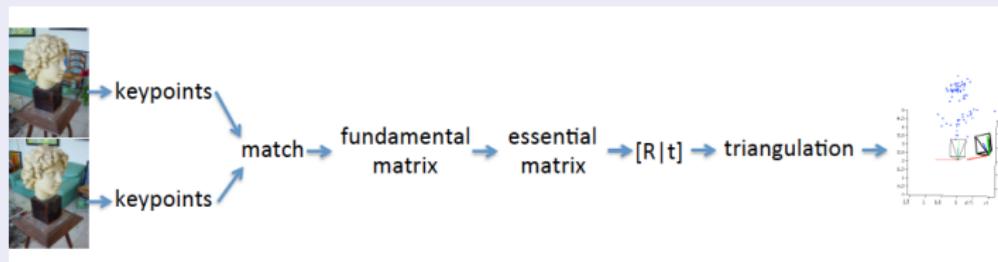
Structure from motion : reconstruction 2D



C'est ce que nous venons de voir !

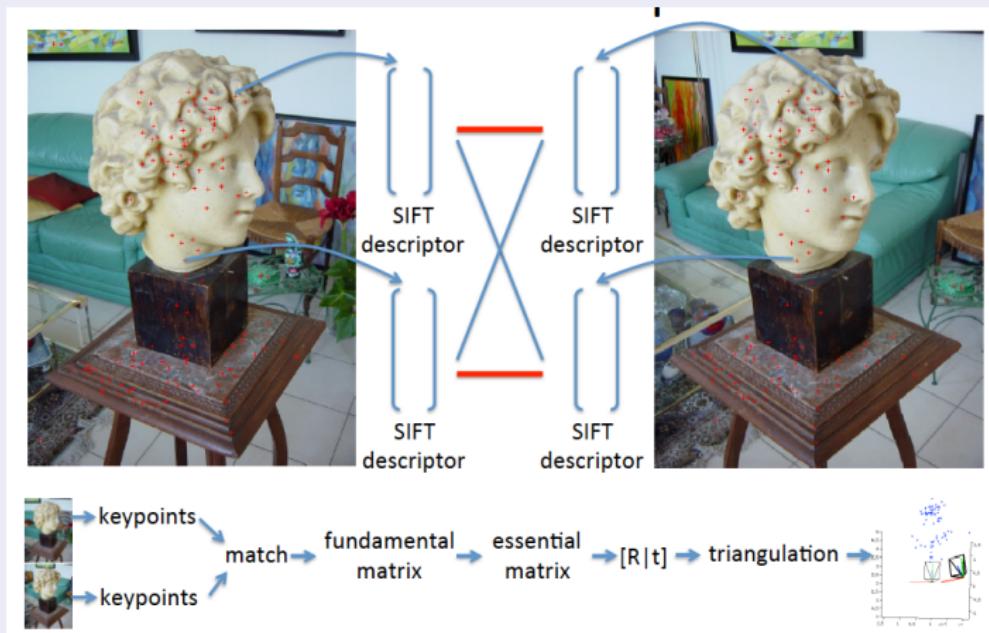
# Reconstruction 3D

## Structure from motion : reconstruction 2D



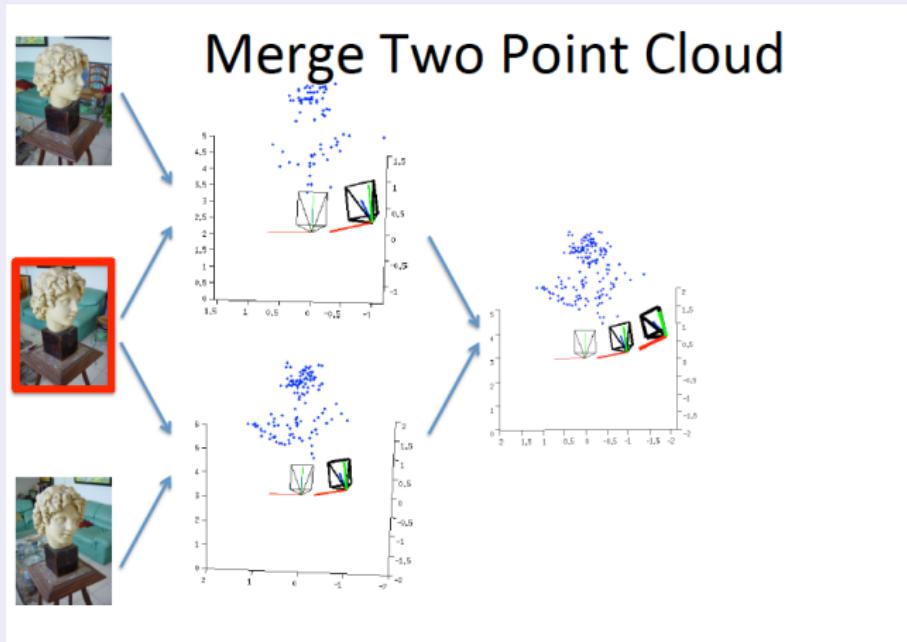
# Reconstruction 3D

## Structure from motion : reconstruction 2D



# Reconstruction 3D

Structure from motion : fusionner les nuages de points.

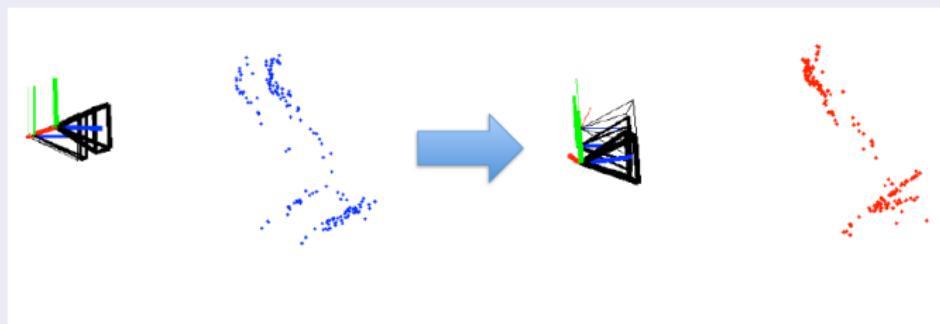


Il ne peut y avoir qu'un seul  $[R|t]$ .

# Reconstruction 3D

Structure from motion : fusionner les nuages de points.

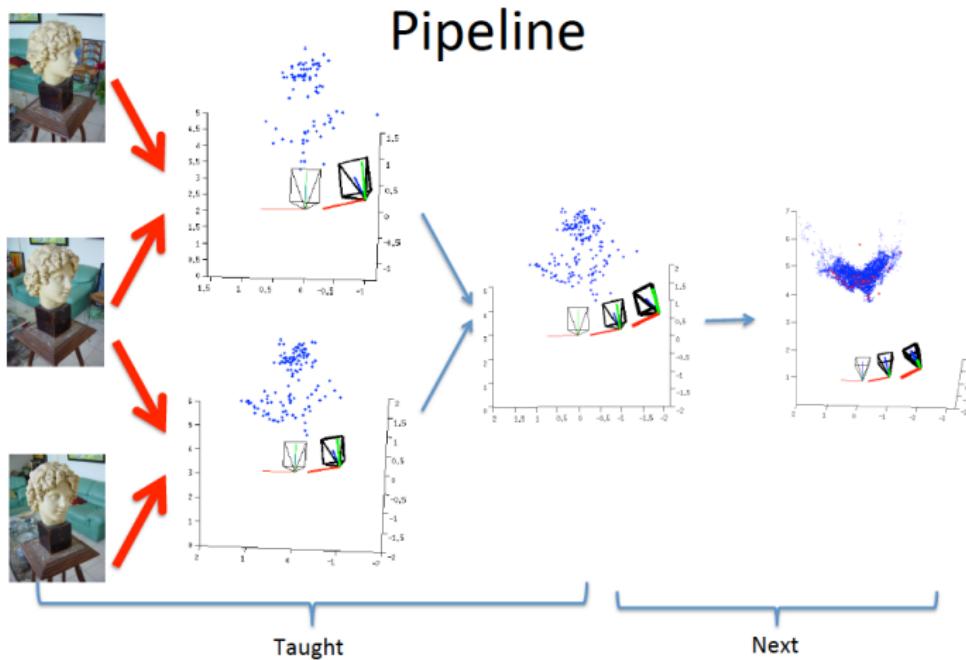
Ajustement de faisceaux (Bundle adjustment)



# Reconstruction 3D

Etape suivante : multiple view stereo : points → plus de points

Idée : propagation des mises en correspondance.



# Structure from motion : applications concrètes

## De nombreuses applications

- Streetview Reconstruction and Recognition :
  - ▶ <http://vision.princeton.edu/projects/2009/TOG/>
  - ▶ <http://vision.princeton.edu/projects/2009/ICCV/>
  - ▶ [research.google.com/pubs/archive/41413.pdf](http://research.google.com/pubs/archive/41413.pdf)
- Photo Tourism : <http://phototour.cs.washington.edu/>
- Microsoft Photosynth : <https://photosynth.net/>
- BigSFM : Reconstructing the World from Internet Photos :  
<http://www.cs.cornell.edu/projects/bigsfm/>
- En robotique : SLAM (Simultaneous Localization and Mapping) :  
<https://www.openslam.org/>

# Plan

- 1 Introduction
- 2 Projection perspective
- 3 Modèle de caméra
- 4 Retour sur la projection perspective
- 5 Calibration d'une caméra
- 6 Triangulation
- 7 Vision stéréoscopique
- 8 Structure from Motion
- 9 Conclusion

# Conclusion

- Les bases de la vision 3D :
  - ▶ Projection perspective
  - ▶ Calibration
  - ▶ Stéréovision
- Beaucoup d'autres travaux pour la reconstruction 3D : géométrie multi-vues, shape from X (shadow, texture, focus), structure from motion, ..
- Beaucoup d'applications actuelles (notamment avec l'apparition de capteur comme la Kinect et cie).  
<http://research.microsoft.com/apps/video/default.aspx?id=152815>
- Explosions du nombre d'images : Building Rome in a day :  
<http://grail.cs.washington.edu/rome/>

# Conclusion

## Pour aller plus loin

- Google Techtalks de S. Seitz :  
<https://www.youtube.com/watch?v=kyIzMr917Rc>
- Tutorial Siggraph :  
<http://www.cs.cmu.edu/~seitz/course/3DPhoto.html>
- The fundamental matrix song :  
<http://danielwedge.com/fmatrix/>