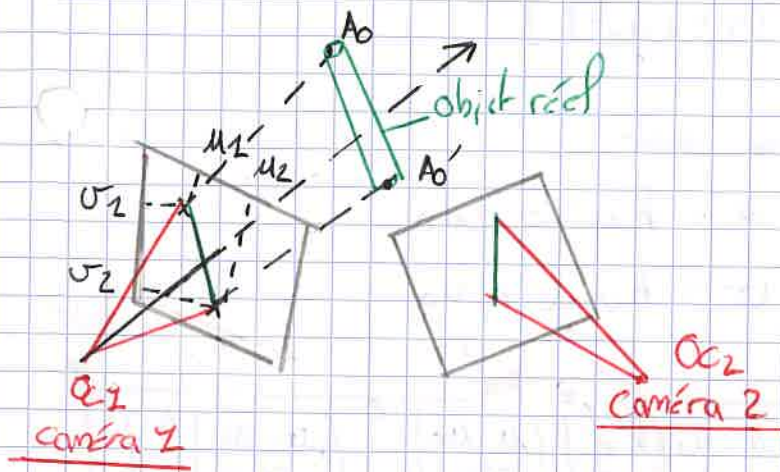


# Cas de la stéréovision :



Problématique: Le repère caméra d'origine  $O_c$  est différent du repère monde (de la pièce)

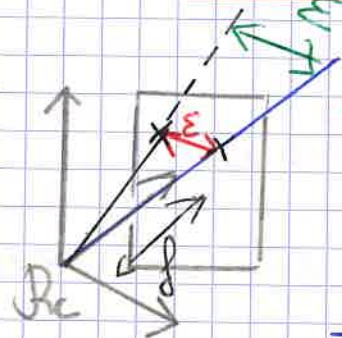
Rappels:  $p_c = \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix}$  décrit par ses paramètres extrinsèques

$$P_c = \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \begin{pmatrix} R_x R_y R_z & T_x \\ 0 & 0 & 0 & T_y \\ 0 & 0 & 0 & T_z \end{pmatrix} \begin{pmatrix} x_m \\ y_m \\ z_m \end{pmatrix}$$

$\begin{cases} \alpha_u = k_u f > 0 & \text{selon les conventions} \\ \alpha_v = k_v f < 0 & \text{selon les conventions} \end{cases}$

Librairie de vision 3D PCL = Point Cloud Library

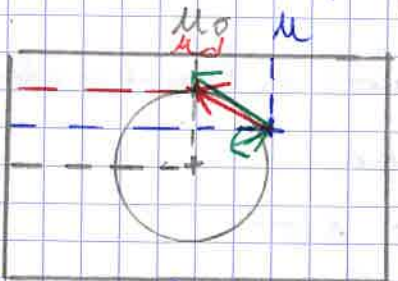
## Modélisation de distortion optique:



⚠ Une erreur sur le plan image  $E$  induit une grande erreur  $\epsilon$  dans la perception de l'espace 3D.  
 => il existe des logiciels de pré-traitement.

L'idée: Utiliser le modèle photogrammétrique

Le déplacement se décompose en un terme radial  $\Delta u_r$  et un terme tangentiel  $\Delta u_t$



Soit un pixel  $P$  de coordonnées  $(u, v)$  s'étant déplacé  $\rightarrow$  par le phénomène de distorsion

5 Le déplacement tangentiel est négligé.



Le modèle est

$$\begin{cases} x_d = x_c [1 + k_2 (x_c^2 + y_c^2)] \\ y_d = y_c [1 + k_2 (x_c^2 + y_c^2)] \end{cases}$$

On sait que la transformation repère - caméra donne

$$\begin{cases} u_d = k_u x_d + u_0 \\ v_d = k_v y_d + v_0 \end{cases} \quad \text{et} \quad \begin{cases} u = k_u x_c + u_0 \\ v = k_v y_c + v_0 \end{cases}$$

Par le calcul on obtient :

En notant : ordre tangentiel  
 $r_2 \in \phi$   
 $r_3 \in \phi$   
 $\uparrow$   
 ordre radial

$$\begin{cases} u_d = u + (u - u_0) k_2 f^2 \left[ \left( \frac{u - u_0}{\alpha_u} \right)^2 + \left( \frac{v - v_0}{\alpha_v} \right)^2 \right] \\ v_d = v + (v - v_0) k_2 f^2 \left[ \left( \frac{u - u_0}{\alpha_u} \right)^2 + \left( \frac{v - v_0}{\alpha_v} \right)^2 \right] \end{cases}$$

$\Delta u$

on a à disposition

$r_2$  to :  $k_2$   
 $r_3$  to :  $k_1; k_2$  et  $k_3$   
 $r_3$  et :  $k_1; k_2; k_3; p_1$  et  $p_2$

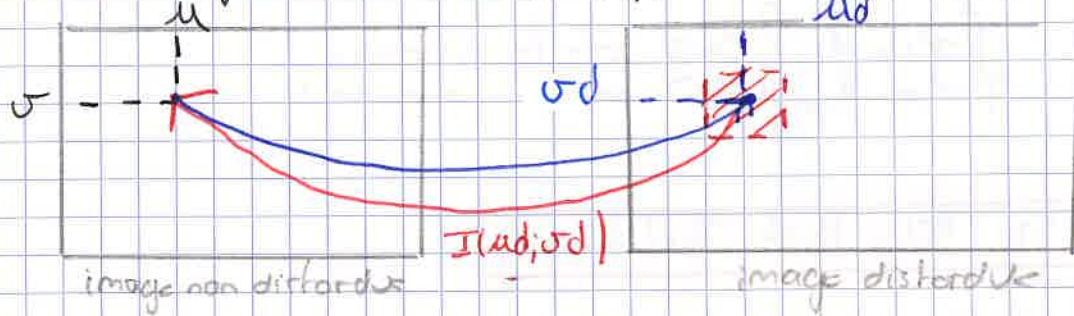
à choisir selon la complexité

du modèle utilisé.

En général si  $f > 7 \text{ mm}$  alors prendre  $r_2$  to

si  $f < 7 \text{ mm}$  alors prendre  $r_3$  to ou  $r_3$  et

Construction de l'image non-distordue : Supposons que l'on connaît les paramètres  $u_0$



On balaye chaque couple  $(u, v)$

on calcule le déplacement associé à chaque pixel

On calcule l'intensité des pixels alentours par la reconstruction sur l'image distordue

Pour des pixels où il y a trop d'erreur alors remplacement par du noir  $\Rightarrow$  zone d'ombre sur les bords de l'image

En pratique, on utilise des tables précalculées en ligne et en colonne par une caméra donnée :

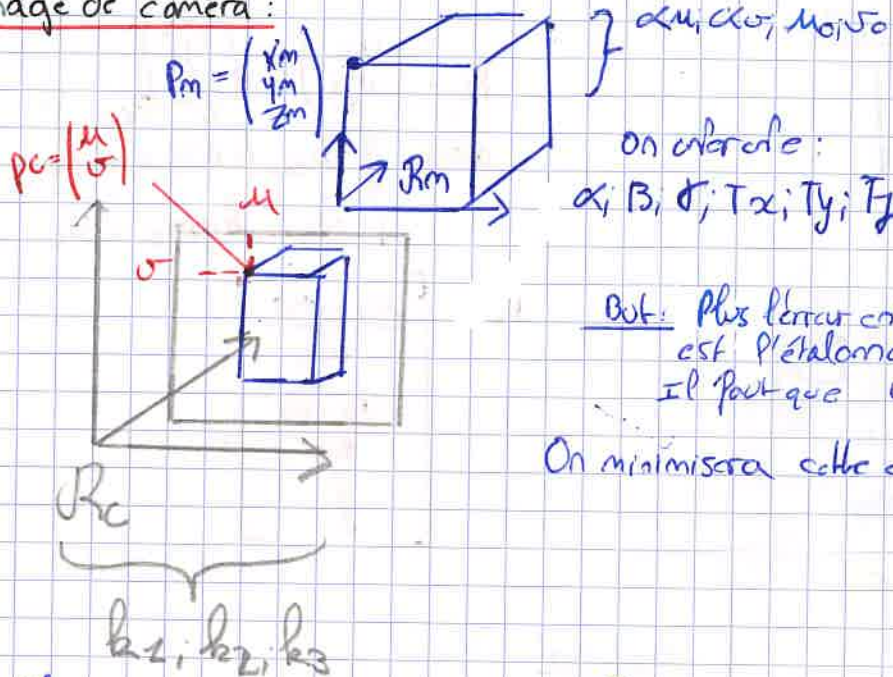
image de correction contenant les  $\Delta u$

image de correction contenant les  $\Delta v$

6  $\Rightarrow$  gain de ressources de calcul



## Étalonnage de caméra :



on cherche :  
 $\alpha, \beta, \sigma; T_x; T_y; T_z$

But : Plus l'erreur entre  $P_m$  et  $P_c$  est faible, meilleur est l'étalonnage.

Il faut que  $u_{u_i}$  et  $v_{v_i} \rightarrow 0$ .

On minimisera cette erreur par les Moindres Carrés

On appelle appariements ZD/3D :  $\{u_i^d; v_i^d; x_m^d; y_m^d; z_i^d\}_{i=1, \dots, m}$

$\Rightarrow$  L'optimisation est non linéaire

$$\begin{cases} u_i + \Delta u_i = P(x) \\ v_i + \Delta v_i = Q(x) \end{cases}$$

Il faut linéariser

$$\begin{cases} P(x) = P(x_0) + \frac{\partial P}{\partial \mu_u} \Delta \mu_u + \dots + \frac{\partial P}{\partial T_x} \Delta T_x + \frac{\partial P}{\partial k_1} \Delta k_1 \\ Q(x) = Q(x_0) + \dots \end{cases}$$

Ainsi

$$\begin{bmatrix} \Delta u_i \\ \Delta v_i \end{bmatrix} = \begin{bmatrix} \frac{\partial P}{\partial \mu_u} & \dots & \frac{\partial P}{\partial k_3} \\ \frac{\partial Q}{\partial \mu_u} & \dots & \frac{\partial Q}{\partial k_3} \end{bmatrix} \begin{bmatrix} \Delta \mu_u \\ \vdots \\ \Delta k_3 \end{bmatrix} - \begin{bmatrix} u_i - P(x_0) \\ \vdots \\ v_i - Q(x_0) \end{bmatrix}$$

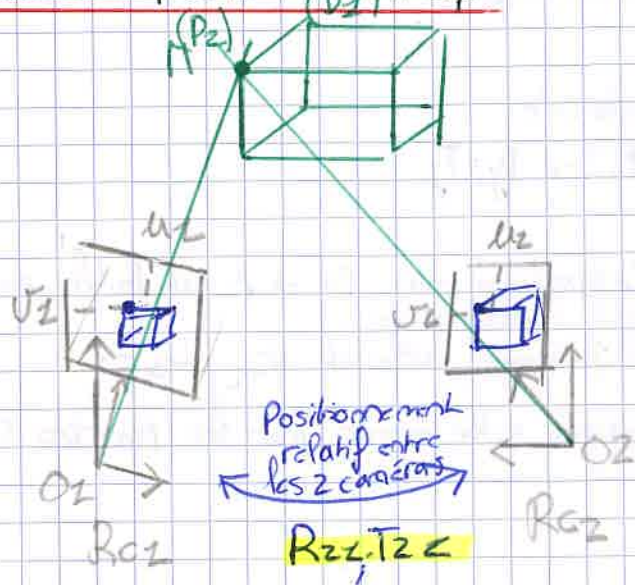
$$\min(V^T V) \Rightarrow \Delta X = \underbrace{(A^T A)^{-1}}_{\text{dit pseudo-inverse}} A^T L$$

Rappel :  $\frac{\partial V^T V}{\partial X} = 0$

Remarque : En général, un bon étalonnage est de l'ordre de  $C(x) = \frac{1}{6}$  pixel



# Stéréovision passive - Principe



$$P_2 = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} R_{02}$$

avec  $\begin{cases} m_2 = k u x_2 + m_0 \\ v_2 = k v y_2 + v_0 \end{cases}$

$$P_2 = \begin{bmatrix} \frac{m_2 - m_0}{k u} \\ \frac{v_2 - v_0}{k v} \\ f \end{bmatrix}$$

$$\vec{Ocp_2} = \begin{bmatrix} \frac{m_2 - m_0}{\alpha u} \\ \frac{v_2 - v_0}{\alpha v} \\ 1 \end{bmatrix} \quad \text{et} \quad \vec{Ocp_2} = \begin{bmatrix} \frac{m_2 - m_0'}{\alpha u'} \\ \frac{v_2 - v_0'}{\alpha v'} \\ 1 \end{bmatrix}$$

$$\begin{cases} \vec{Ocp_2} = 1 \vec{Ocp_2} & (M \in D_2) \text{ dans } R_{02} \\ \vec{Ocp_2} = 1 \vec{Ocp_2} & (M \in D_2) \text{ dans } R_{02} \end{cases} \quad \updownarrow$$

$R_{22} \vec{Ocp_2}$

$$\begin{pmatrix} x_{11} - t_{22x} \\ y_{11} - t_{22y} \\ z_{11} - t_{22z} \end{pmatrix} = 1 \vec{R_{22} \vec{Ocp_2}}$$

En effet,  $\begin{pmatrix} t_{22x} \\ t_{22y} \\ t_{22z} \end{pmatrix} = \begin{pmatrix} - & - & - & | & t_{22x} \\ 1 & R_{22} & | & t_{22y} \\ 0 & 0 & 0 & | & t_{22z} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

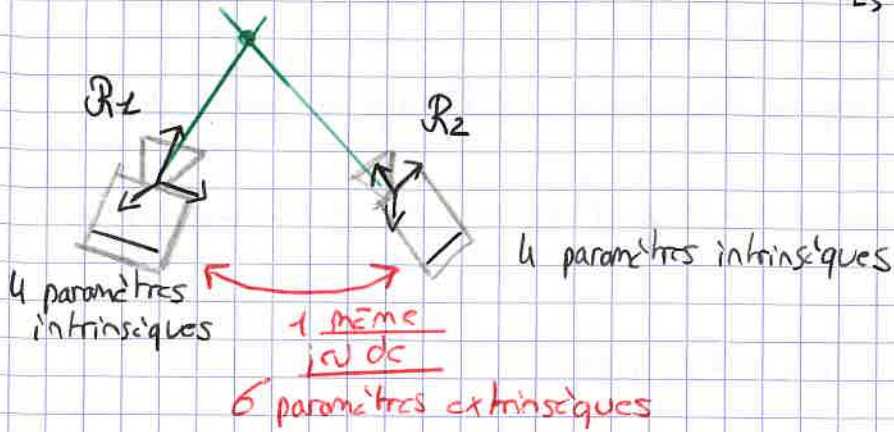
↓  
Coordonnées de  $\vec{Ocp_2}$  dans  $R_{02}$

Remarque: La position relative entre les caméras nous est donnée par les paramètres extrinsèques issus de leurs étalonnages respectifs. ⚠ L'objet étalon doit être dans la même position



Rappels : Stéréovision :

→ Vergence : angle entre les axes optiques  
↳ dépend de l'éloignement de la scène



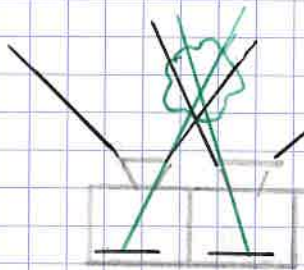
Positionnement d'un caméra par rapport à l'autre :

Si caméras accolées :

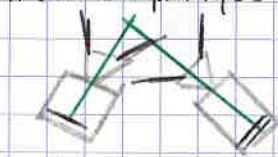
- avantage : champ commun grand
- inconvénient : problème de précision

Si caméras éloignées :

- avantage : plus de précision
- inconvénient : champ de vue commun moins important



→ solution la plus utilisée en pratique



stéréovision = caméra proches spatialement



vision binoculaire

⇒ précis mais champ de vue très différent ne respecte pas l'hypothèse de la stéréovision

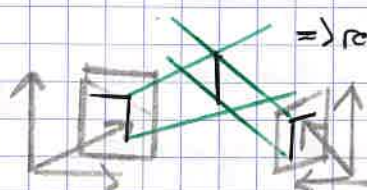
! Précompte d'occlusion : un objet cache une autre partie de la scène que l'on souhaiterait voir avec la position des caméras.  
⇒ a priori on ne pourra pas reconstruire toute la scène.

On reconstruit ce qui est commun = apparier les pixels communs

Les images peuvent être prises de manière synchrone (majorité des cas) ou non (ex : caméra unique montée sur l'OT d'un bras robotisé qui prend deux points de vue d'une scène immobile)

Stéréovision passive :

Appariement de segments



⇒ reconstruction de segment 3D  
utilisé pour les scènes avec des constructions numériques



Corrélation = appariement d'une de pixels

Appariement de pixels: utilisation de la ZNCC:

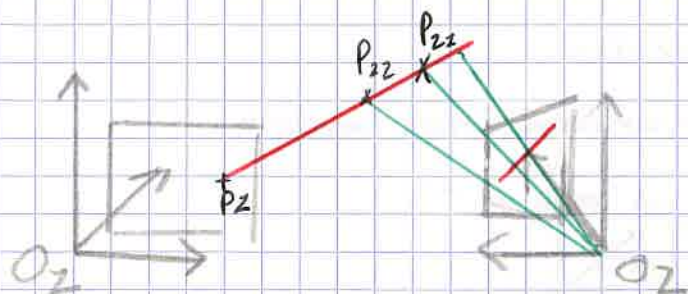
$\bar{I}$  et  $\bar{I}'$  sont respectivement la moyenne des niveaux dans l'image de gauche et de droites.

choix d'un seul pour valider l'appariement du pixel  $p_z$  (0,6-0,7 en extérieur et 0,9 en intérieur pour le score obtenu)

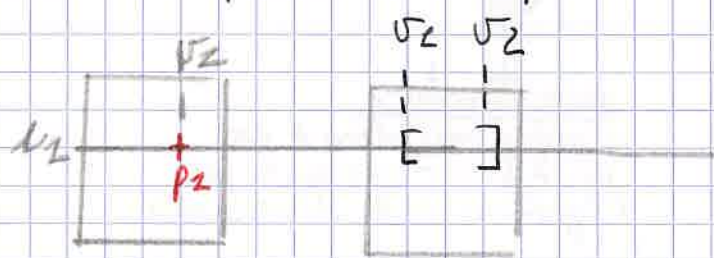
Problématique: Comment réduire le coût de calcul?

Contrainte épipolaire: On se ramène à un balayage de la droite

l'ensemble des points  $p_z$  candidats appartient à la droite épipolaire c.à.d. à l'intersection des plans images ~~appariement~~ contenant ~~appariement~~ ( $p_z$ ,  $O_z$ ,  $O_z'$ )



De plus, il suffit que les axes optiques soient coplanaires pour que les pixels partagent la même coordonnées  $u$  (réaliser par un algorithme en pratique ou mécaniquement dans de rares cas par ex au CUES)



$$|u_z - u_z'| = \text{disparité}$$

$\Rightarrow$  recherche limitée à quelques pixels dans une intervalle

Enfin,

La profondeur  $Z = \frac{K}{|u_z - u_z'|}$  la dimension que l'on souhaite connaître avec  $K$  constante

Reconstruction:

Hypothèses:  
• Les axes optiques sont coplanaires  
• Les axes optiques sont parallèles

$$\begin{cases} x_z = f \frac{x_z}{z_z} \\ y_z = f \frac{y_z}{z_z} \end{cases}$$

$$\text{et } \begin{cases} u_z = h_u x_z + u_0 \\ v_z = h_v y_z + v_0 \end{cases}$$

de même pour la caméra 2



$$y_z = y_z + b \Rightarrow \frac{y_z z_z}{f} = \frac{y_z z_z}{f'} + b$$

$$z_1 = z_2 \Rightarrow z \left( \frac{y_z}{f} - \frac{y_z}{f'} \right) = b \quad \text{or} \quad y_z = \frac{v_z - v_0}{k_v}$$

$$\left\{ \begin{array}{l} y_z = \frac{v_z - v_0'}{k_v'} \end{array} \right.$$

Ainsi,  $z = \frac{b}{\left| \frac{v_1 - v_0}{\alpha_v} - \frac{v_2 - v_0}{\alpha_v'} \right|}$

si les paramètres intrinsèques sont identiques c.à.d.  $\left\{ \begin{array}{l} \alpha_v = \alpha_v' \\ v_0 = v_0' \end{array} \right.$

alors  $z_1 = z_2 = z = \frac{b \alpha_v}{|v_1 - v_2|}$  on pose  $d = |v_1 - v_2|$

Les autres coordonnées sont  $u_z = k_u \int \frac{x_z}{z_z} + u_0 \Rightarrow x_z = z \left( \frac{u_z - u_0}{\alpha_u} \right)$

$$v_z = k_v \int \frac{y_z}{z_z} + v_0 \Rightarrow y_z = z \left( \frac{v_z - v_0}{\alpha_v} \right)$$

Le point dans la scène réelle est à la position  $(x_z; y_z; z_1 = z)$

Remarque : En pratique, on s'arrange pour modifier les paramètres intrinsèques d'une caméra pour qu'ils soient identiques

- La disparité a une unité parfois subpixelique (on a une précision qui descend en dessous du pixel)
- On peut obtenir l'image de disparité  $|v_z - v_z|$ 
  - des valeurs homogènes signifie que la scène est uniforme
  - des valeurs très éloignées montre une erreur d'appariement → possible de la filtrer avant reconstruction

Application robotique mobile:

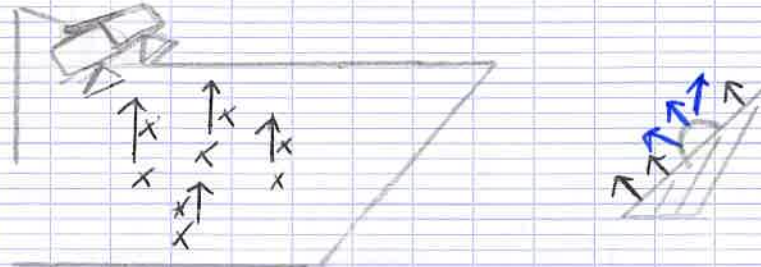
2 têtes stéréos embarquées

- Évitement immédiat
- planification

Pour déterminer si un terrain est praticable on calcule le produit vectoriel



de deux points proches On obtient ainsi la normale au plan. On a donc un nuage de normale  $\Rightarrow$  une irrégularité dans la direction de ces vecteurs signale un obstacle :



Représentation  
 $2D^{1/2}$  ( $z$  est fonction de  $x$  et  $y$ )

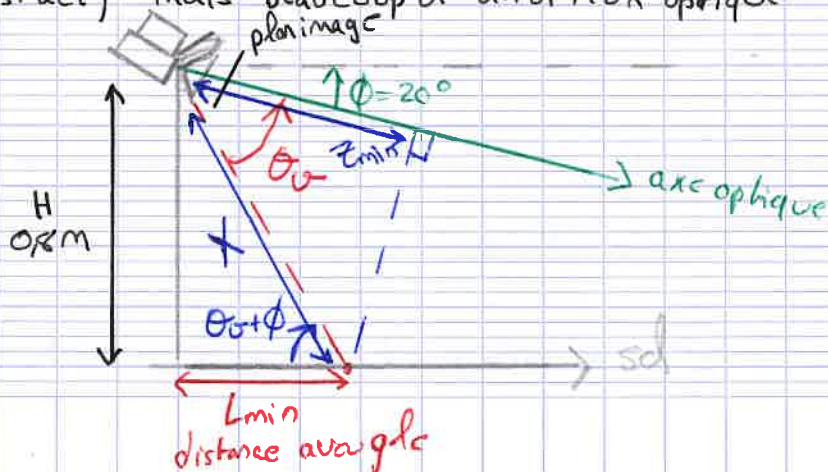
Fusion de plusieurs vue : une prise de vue limite notre perception de la scène  $\Rightarrow$  on déplace la caméra / banc stéréo et on fusionne les reconstructions = modélisation incrémentale

Refacéliser : plaquer les niveaux de capture sur un modèle 3D pour recrer une vision de l'objet

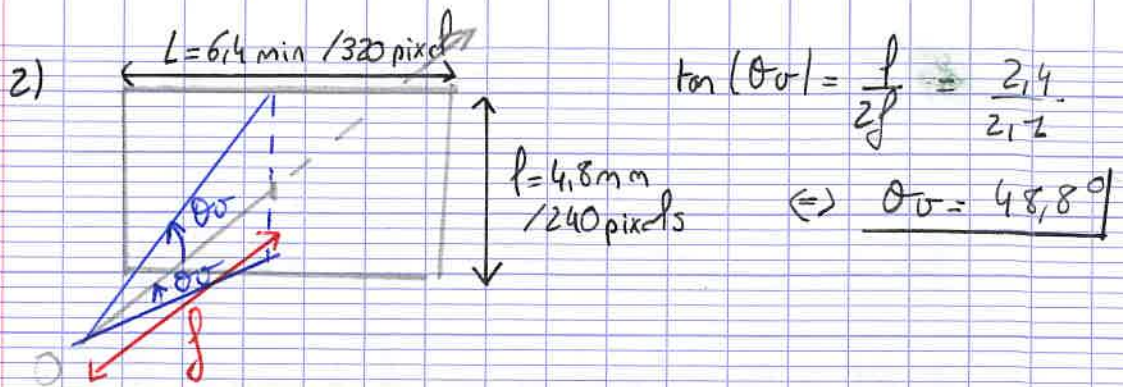
Pour s'affranchir des problèmes de texture uniforme : des lumières proches de l'infrarouge créent artificiellement des textures pour permettre la reconstruction (plutôt utiliser en application intérieure)

Exercice (diapo 26)

- 1) 2,2 mm de focale  $\Rightarrow$  grand angle (pertinent pour la détection d'obstacle) mais beaucoup de distorsion optique







$$\tan(\theta_v + \phi) = \frac{H}{L_{\min}} \Leftrightarrow L_{\min} = \frac{H}{\tan(\theta_v + \phi)} = 0.3 \text{ m}$$

3)  $\alpha_v = k_v \cdot f$

$k_v = \frac{320}{6.4} \Rightarrow \alpha_v = \frac{320 \times 2.1}{6.4} \approx 105$

$\rightarrow \text{en [pix/mm]}$

distance  
base inter-caméras

Rappel:  $Z = \frac{B \alpha_v}{|u_z - v_z|}$  ← des paramètres intrinsèques

disparité disp

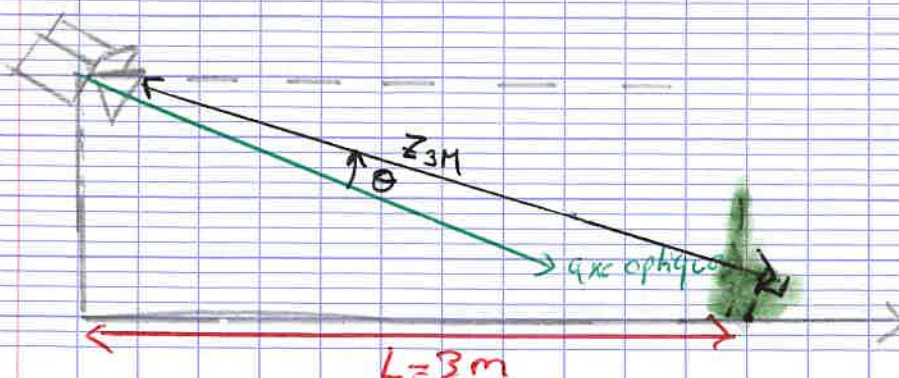
$$\cos(\theta_v) = \frac{Z_{\min}}{X}$$

$$\cos(\theta_v + \phi) = \frac{L_{\min}}{X}$$

$$Z_{\min} = L_{\min} \frac{\cos(\theta_v)}{\cos(\theta_v + \phi)} \quad (\text{coordonnée portée par l'axe optique})$$

$$\text{disp}_{\max} = \frac{B \alpha_v}{Z_{\min}} \times \frac{\cos(\theta_v + \phi)}{\cos(\theta_v)} = 22 \text{ pixels}$$

Si on éloigne le point d'observation alors  $\theta_v$  risque de passer au dessus de l'axe optique. on va signer les angles dans le sens trigonométrique





Pour  $L = 3m$  ;  $\tan(\theta + \phi) = \frac{H}{L} \Rightarrow \theta = +5^\circ$   
 $\Rightarrow \text{disp} = 4 \text{ pixels}$

Pour  $L = 5m$  ;  $\theta = 71^\circ$  et  $\text{disp} = 2 \text{ pixels}$

Au delà de 5m, il n'est pas possible de mesurer la disparité. Pour une application d'aide au stationnement, cela est suffisant.

On construit d'abord une image de référence de disparité sur un sol plat.

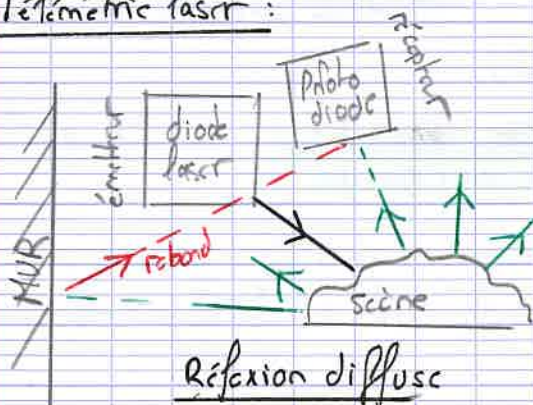
En utilisation, on réalise la différence entre cette image de référence et une image à traiter (potentiellement avec un obstacle) si les valeurs diffèrent alors présence d'un obstacle.

Cela ne peut pas être fait avec des images classiques parce que la disparité est indépendante de la ges. luminosité, colorimétrie, etc. de la scène. La disparité représente une géométrie

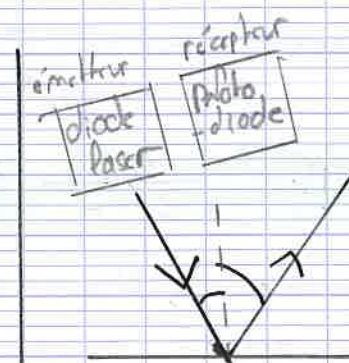
## Capturs Actifs

20/09/2022

### Télémétrie laser :



⚠ La mesure peut être ébranlée par le rebond



Réflexion spéculaire  
 $\Rightarrow$  quasiment pas de chance de récupérer l'écho

On obtient des images de profondeurs :

