

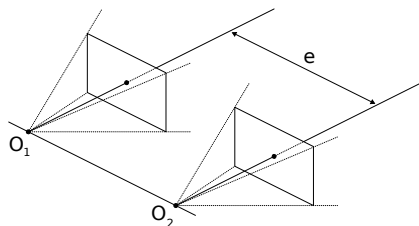
Rectification épipolaire stéréovision dense

François Cabestaing

Master ASE et Informatique, spécialité "Image Vision Interaction"

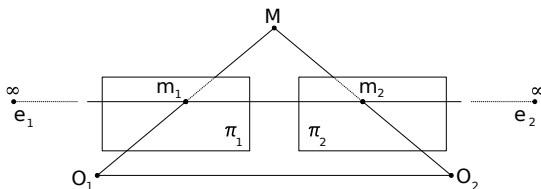
configuration canonique d'un stéréoscope

Configuration la plus souvent utilisée en pratique, du fait que la mise en correspondance stéréoscopique est simplifiée.



- les deux caméras sont identiques (paramètres intrinsèques).
- les axes optiques sont parallèles, séparés par un *entraxe* e .
- les deux capteurs d'image sont coplanaires.
- les lignes horizontales des capteurs sont confondues.

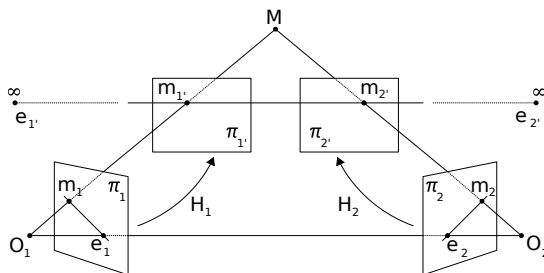
propriétés épipolaires de la configuration canonique



- les plans projectifs sont confondus, mais on conserve généralement deux notations, car les repères associés sont différents.
- les droites épipolaires associées à deux points homologues sont confondues.
- les épipoles sont à l'infini, dans une direction définie par la droite de base.

rectification épipolaire

Objectif : revenir à la configuration canonique quand le stéréoscope est quelconque



- principe : transformer les deux images acquises par un stéréoscope non canonique par des homographies.
- estimation des homographies qui permettent de conserver la plus grande partie commune sur les images.

détermination des homographies

La matrice fondamentale d'un stéréoscope en configuration canonique est :

$$\bar{\mathbf{F}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}^{\times} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

Les homographies qui permettent la rectification épipolaire donnent :

$$\mathbf{m}_{1'} = \mathcal{H}_1 \mathbf{m}_1 \quad \text{et} \quad \mathbf{m}_{2'} = \mathcal{H}_2 \mathbf{m}_2$$

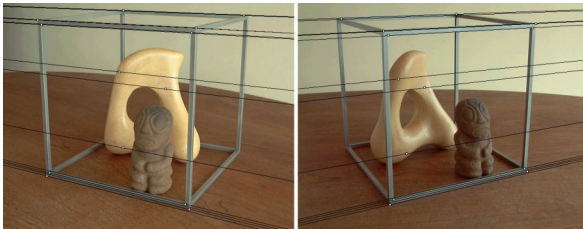
Si \mathbf{m}_1 et \mathbf{m}_2 sont homologues, alors $\mathbf{m}_{1'}$ et $\mathbf{m}_{2'}$ également et donc :

$$\mathbf{m}_{2'}^T \bar{\mathbf{F}} \mathbf{m}_{1'} = \mathbf{m}_2^T \mathcal{H}_2^T \bar{\mathbf{F}} \mathcal{H}_1 \mathbf{m}_1 = 0 = \mathbf{m}_2^T \mathbf{F} \mathbf{m}_1$$

Donc les homographies \mathcal{H}_1 et \mathcal{H}_2 doivent vérifier :

$$\mathbf{F} = \mathcal{H}_2^T \bar{\mathbf{F}} \mathcal{H}_1$$

rectification épipolaire évitant les déformations (Loop & Zhang)



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



avec déformation

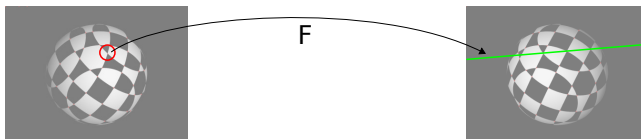


déformation minimale

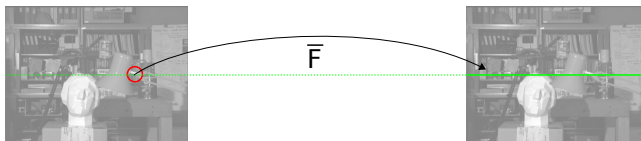
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.5603>

contrainte épipolaire

Le point homologue \mathbf{m}_r dans l'image droite d'un point \mathbf{m}_l de l'image gauche est situé sur la droite épipolaire de ce dernier.



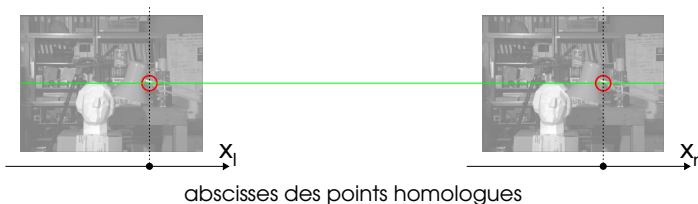
configuration quelconque



configuration canonique

disparité

En configuration canonique, la disparité est la différence entre les abscisses de deux points homologues.



Si l'image de gauche est prise comme référence, la disparité est donnée par :

$$d_l(\mathbf{m}_l) = d_l((x_l \ y)^T) = x_l - x_r$$

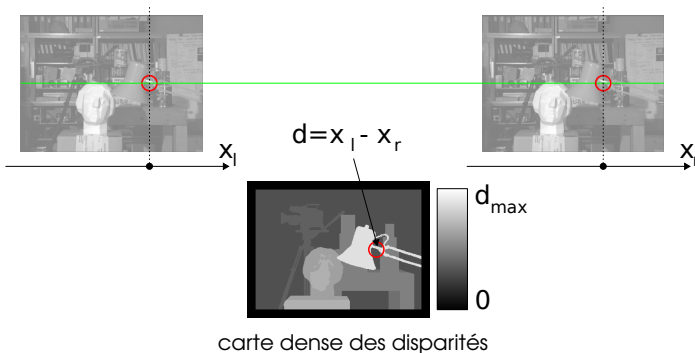
Si l'image de droite est prise comme référence, la disparité est donnée par :

$$d_r(\mathbf{m}_r) = d_r((x_r \ y)^T) = x_l - x_r$$

carte des disparités

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

- carte éparse : une disparité est associée à quelques pixels
- carte dense : une disparité est associée à chaque pixel



autres contraintes pour la recherche des homologues

- contrainte d'unicité

à chaque pixel de l'image de référence est associé au maximum un pixel de l'autre image, et inversement.

- contrainte d'ordre (pas toujours vérifiée)

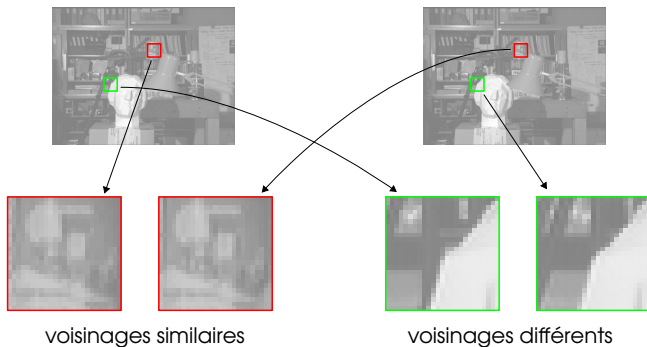
quand $(\mathbf{m}_{l1}, \mathbf{m}_{r1})$ et $(\mathbf{m}_{l2}, \mathbf{m}_{r2})$ sont deux paires de points homologues situés sur la même ligne, alors :

$$\text{si } x_{l1} < x_{l2}, \text{ alors } x_{r1} < x_{r2}$$

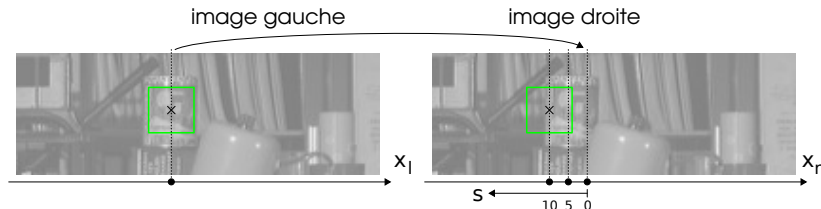
- contraintes liées à la photométrie

similarité des niveaux de gris dans les voisinages

- condition : surfaces lambertiennes
- contrainte non vérifiée à proximité des discontinuités



maximisation de la similarité / minimisation de la dissimilarité



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

fonctions de dissimilarité

$I_l(x_l, y_l)$ est le niveau de gris du pixel (x_l, y_l) dans l'image de gauche.

$I_r(x_r, y_r)$ est le niveau de gris du pixel (x_r, y_r) dans l'image de droite.

- SAD : sum of absolute differences

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

- SSD : sum of squared differences

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

dans les deux cas, le voisinage est de dimension $(2w_x + 1) \times (2w_y + 1)$.

Problème : nécessite que les réponses photométriques des deux caméras soient identiques.

fonction de similarité basée sur une corrélation

Insensible aux différences entre les réponses photométriques des deux caméras.

- NCC : normalized cross-correlation

$$\text{NCC}(x_l, y, s) =$$

$$\frac{\sum_{i=-w_x}^{w_x} \sum_{j=-w_y}^{w_y} \left(I_l(x_l + i, y + j) - \bar{I}_l(x_l, y) \right) \left(I_r(x_l + i - s, y + j) - \bar{I}_r(x_l - s, y) \right)}{\sqrt{\sum_{i=-w_x}^{w_x} \sum_{j=-w_y}^{w_y} I_l(x_l + i, y + j)^2} \cdot \sqrt{\sum_{i=-w_x}^{w_x} \sum_{j=-w_y}^{w_y} I_r(x_l + i - s, y + j)^2}}$$

dans laquelle $\bar{I}_l(x_l, y)$ et $\bar{I}_r(x_l - s, y)$ sont les moyennes locales du niveau de gris dans un voisinage :

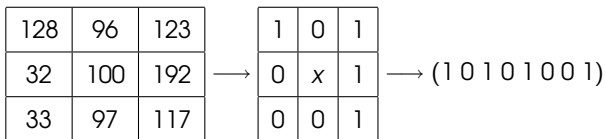
$$\bar{I}(x, y) = \frac{1}{(2w_x + 1) \cdot (2w_y + 1)} \sum_{i=-w_x}^{w_x} \sum_{j=-w_y}^{w_y} I(x + i, y + j)$$

Inconvénient : temps de calcul élevé.

fonction de dissimilarité non-métrique : census + Hamming

transformée census : associe 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
- un bit de la chaîne vaut 0 si le niveau de gris du pixel correspondant est inférieur à celui du pixel central



distance de Hamming : nombre de bits différents dans deux chaînes

exemple : $d_H((1\ 0\ 1\ 0\ 1\ 0\ 0\ 1), (1\ 0\ 0\ 1\ 1\ 0\ 1\ 1)) = 3$

combinaisons de contraintes, puis optimisation

fonction de coût = combinaison des contraintes

- contrainte épipolaire
- contraintes d'ordre + contrainte d'unicité
- contrainte de similarité des voisinages

approches locales

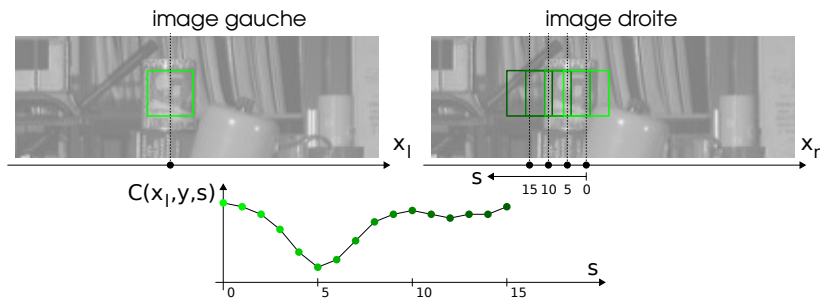
L'extremum de la fonction est recherché pour chaque pixel indépendamment des voisins.

approches globales

L'extremum de la fonction est recherché en tenant compte des voisins du pixel traité.

Le voisinage peut être l'image complète.

approche locale : winner takes all (WTA)



- on fixe un décalage maximal s_{max}
- on calcule la fonction de coût $C(x_l, y, s)$ pour chaque décalage s compris entre 0 et s_{max} .
- on détermine l'extremum (maximum ou minimum) de la fonction

$$\hat{d}_l(x_l, y) = \arg \min_s C(x_l, y, s)$$

exemple d'approche globale : chemin optimal

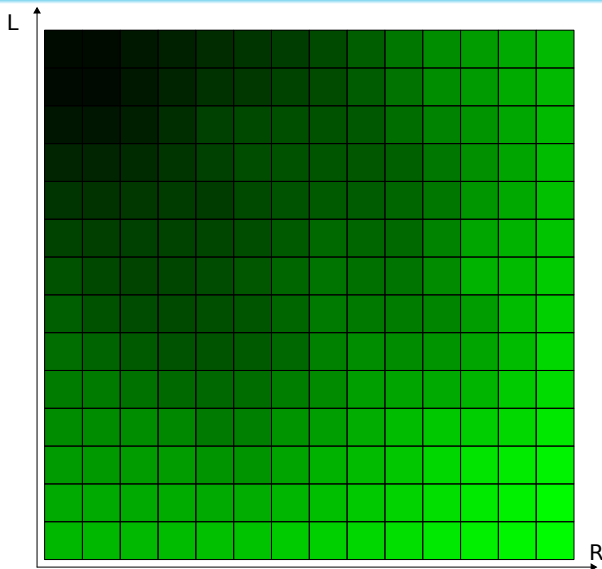
prise en compte des similarités sur une ligne complète de taille N_x .

- tableau 2D de taille $N_x \times N_x$
- une case de coordonnées (x_l, x_r) contient la valeur de la fonction de similarité $C(x_l, y, x_l - x_r)$

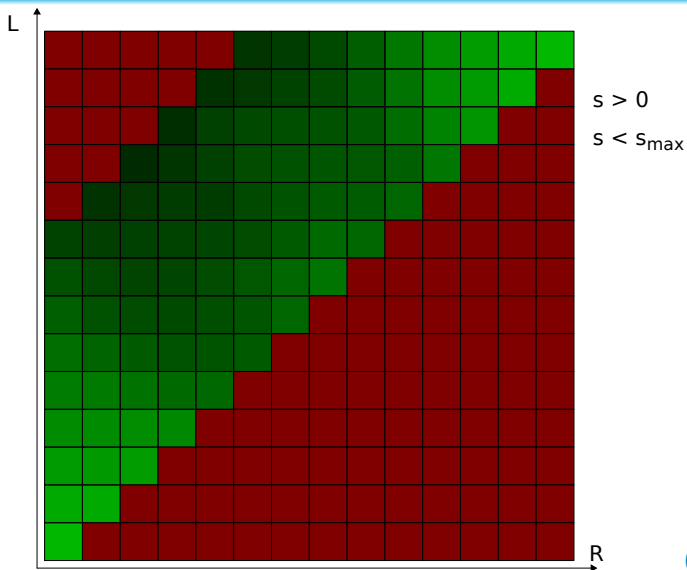
prise en compte de contraintes globales.

- la fonction globale de coût évalue toutes les correspondances possibles des pixels de la ligne
- correspond à un chemin dans le tableau des similarités
- termes de pénalité défavorisant les chemins avec beaucoup de discontinuités

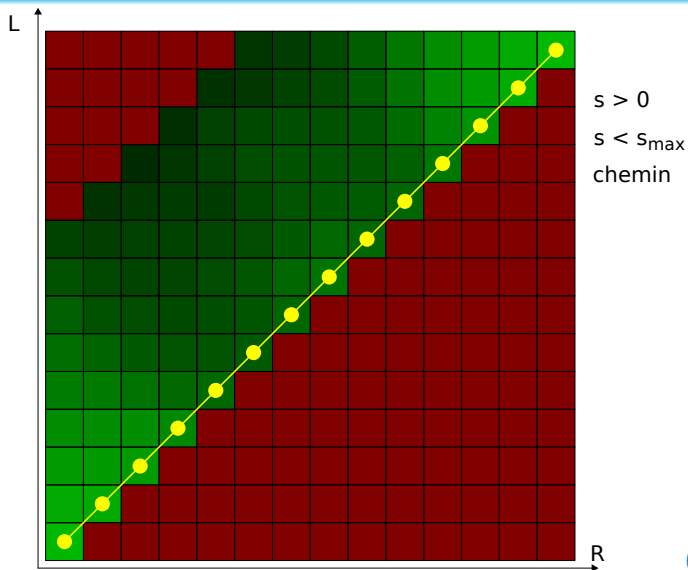
chemin optimal, exemple (1/4)



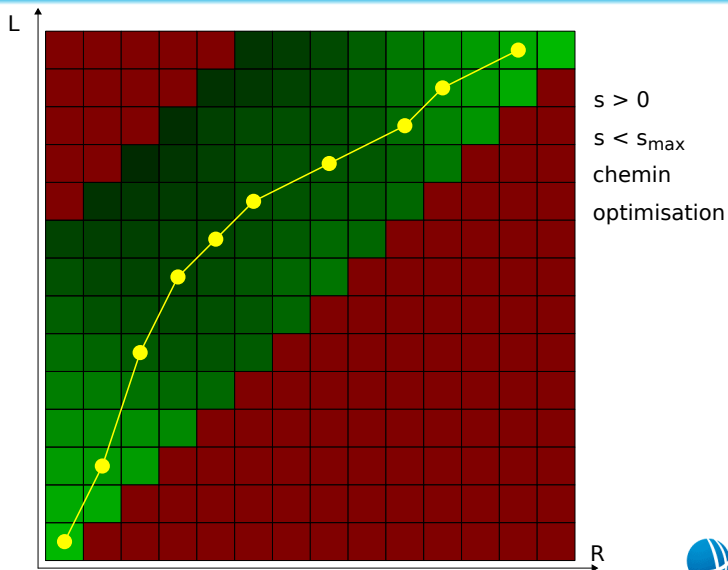
chemin optimal, exemple (2/4)



chemin optimal, exemple (3/4)



chemin optimal, exemple (4/4)

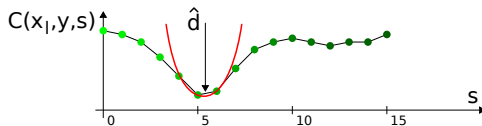


amélioration sub-pixel

Objectif : obtenir une estimation de la disparité avec une précision inférieure au pixel

- interpolation sur plusieurs voisins
- modélisation locale de la fonction de coût

exemple : approche WTA sur coût local



modèle parabolique à proximité du minimum

cohérence gauche-droite

Objectif : vérifier que les paires de pixels homologues estimées restent identiques lorsqu'on change l'image de référence

Avec l'image gauche comme référence, la coordonnée du pixel droit vaut :

$$x_r = x_l - \hat{d}_l(x_l, y)$$

Avec l'image droite comme référence, la coordonnée du pixel gauche vaut :

$$x_l = x_r + \hat{d}_r(x_r, y)$$

Donc les disparités estimées doivent vérifier :

$$\hat{d}_r(x_r, y) = \hat{d}_l(x_r + \hat{d}_r(x_r, y), y)$$

$$\hat{d}_l(x_l, y) = \hat{d}_r(x_l - \hat{d}_l(x_l, y), y)$$

Permet d'éliminer les *faux appariements* causés par les *occultations*.

TP : minimisation du SSD

- calcul du SSD avec l'image gauche comme référence
- idem sur image droite
- vérification de cohérence
- calcul rapide des indices de similarité SSD

bibliographie

Éléments de cours

- Frédéric Devernay, INRIA Grenoble, cours "Vision par ordinateur 3-D".
<http://devernay.free.fr/cours/vision/>
- Cours de Sébastien Roy sur la vision 3D, Université de Montréal.
http://www.iro.umontreal.ca/~roys/fr_ift6145H07.shtml

Rapport de recherche INRIA

- O. Faugeras, B. Hotz, H. Mathieu, *et coll.*, *Real time correlation-based stereo : algorithm, implementations and applications*, RR-2013 INRIA (2003). <http://hal.inria.fr/docs/00/07/46/58/PDF/RR-2013.pdf>

Articles de chercheurs

- D. Scharstein and R. Szeliski : *A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms*. International Journal of Computer Vision 47(1-3) : 7-42 (2002)