

Filtre a contrario et raffinement sous-pixellique pour la chaîne stéréo

Neus Sabater, Pascal Monasse, Jean-Michel Morel

18 novembre 2011

Plan

- 1 Introduction
- 2 Élimination de fausses correspondances
 - Le match de blocs a contrario
 - Le filtre d'autosimilarité
- 3 Précision sous-pixellique
- 4 Projets

Contrainte épipolaire

- On dispose de deux vues de la même scène.
- On se place dans le repère lié à la caméra 1 :

$$\lambda x = KX \quad \lambda' x' = K'(RX + T)$$

- On réécrit sous forme de système :

$$\begin{pmatrix} I_3 & -K^{-1}x & 0 & 0 \\ R & 0 & K'^{-1}x' & T \end{pmatrix} \begin{pmatrix} X \\ \lambda \\ \lambda' \\ 1 \end{pmatrix} = 0$$

- Le déterminant de cette matrice 6×6 doit donc être nul :

contrainte épipolaire

$x'^T F x = 0$ avec $F = K'^{-T} [T]_{\times} R K^{-1}$, matrice fondamentale 3×3

Rectification épipolaire

- x (resp. x') étant donné, x' (resp. x) est sur la droite d'équation Fx (resp. $F^T x'$)
- Toute matrice de rang 2 s'écrit comme matrice F , donc de 7+ points (x_i, x'_i) connus, $\det F = 0$ et $\|F\| = 1$, on peut calculer F , sans avoir besoin des calibrations K et K'
- **Cas particulier** : $K = K'$, $R = I$, $T = \lambda e_1$ (pure translation suivant l'axe des x), $F = [e_1]_{\times}$
- Une fois F calculée, on peut toujours se ramener à une telle situation en appliquant des homographies H et H' aux images de sorte que

$$F = H'^T [e_1]_{\times} H$$

- 2*8 inconnues dans H et H' , 3*3 équations, donc 7 degrés de liberté, d'où de multiples méthodes de rectification épipolaire minimisant la distorsion projective appliquée aux images

Rectification épipolaire



Image 1

Rectification épipolaire



Image 2

Rectification épipolaire



Image 1



Image 1 rectifiée

Rectification épipolaire



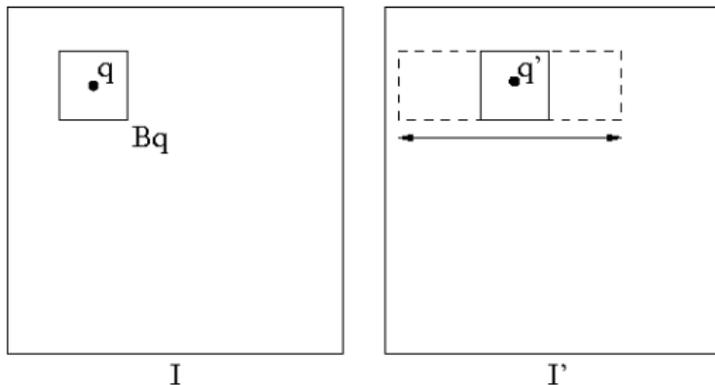
Image 2



Image 2 rectifiée

Stéréoscopie et mise en correspondance des images

- On suppose que les images ont préalablement été rectifiées.
- Le problème est ramené à une dimension : les points homologues se trouvent sur la même droite horizontale.



- On appelle **disparité** au point q son déplacement apparent

$$d(q) = q'_x - q_x$$

Stéréoscopie et mise en correspondance des images

- Reprenons nos équations dans le cas $R = I$ et $T = Be_1$:

$$\lambda x = KX \quad \lambda' x' = K'(X + Be_1)$$

- Supposant $K_{33} = K'_{33} = 1$ et $x_3 = x'_3 = 1$, on a $\lambda = \lambda' = z$ et

$$z(x' - K'K^{-1}x) \cdot e_1 = fB \quad (f = K'_{33})$$

Si $K = K'$, on a la formule :

$$z = Bf/d$$

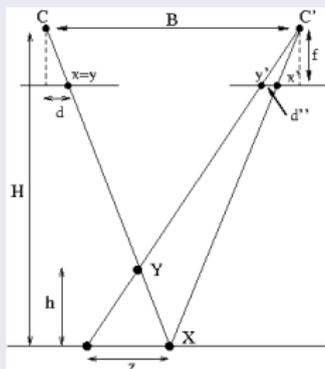
z =profondeur, B =baseline, f = focale (en pixels), d =disparité

La rectification épipolaire n'impose pas $K = K'$ (mais seulement que les deux dernières lignes coïncident), donc **sans calibration** :

- **on peut** rectifier les images et calculer une carte de disparités ;
- par contre, **on ne peut pas** en déduire les profondeurs.

Stéréoscopie et mise en correspondance des images

Principe fondamental de la vision stéréoscopique



$$h \simeq \frac{z}{B/H}, \quad z = d'' \frac{H}{f}.$$

f focale de la caméra.

H distance centre optique-sol.

B distance entre les centres optiques (baseline).

Enjeu

Etant donné deux images rectifiées, la mise en correspondance de points homologues et le calcul du décalage (disparité) entre eux donne l'information sur le relief relatif de la scène.

Approches pour la mise en correspondance

Approches globales : Minimisation d'une énergie de la forme

$$\min_d E_{\text{données}}(d) + \alpha E_{\text{régul}}(d).$$

- Graph-Cuts [Kolmogorov *et al.* 05]
- Programmation dynamique [Ohta et Kanade 85] [Wang *et al.* 06]

Avantages : habituellement pas de phénomène d'adhérence, peu sensibles aux effets stroboscopiques, peu sensibles au bruit, résultats denses.

Inconvénient : nécessite un modèle de régularité valable globalement dans toute l'image.

Approches pour la mise en correspondance

Approches locales :

- *Feature-matching* : mise en correspondance de segments [Schmid et Zisserman 00], lignes de niveau [Musé *et al.* 06], coins [Cao 04], descripteurs SIFT [Lowe 04] [Rabin *et al.* 08], etc. → résultats très épars.
- *Block-matching* : estimation de la disparité ε en chaque point \mathbf{x}_0 en minimisant

$$\mu(\mathbf{x}_0) = \arg \min_d C(\mathbf{x}_0, d)$$

où C est une fonction de coût (NCC, SAD, SSD) calculée en chaque voisinage de \mathbf{x}_0 .

Avantages : règles statistiques sur la **fiabilité** des appariements et la **précision** d'un appariement peut être caractérisée en fonction du bruit.

Inconvénient : phénomène d'adhérence.

Mesures de similitude entre patches

- **SAD** : Sum of Absolute Differences

$$d(q) = \arg \min_d \sum_{p \in F} |I(q + p) - I'(q + de_1 + p)|$$

- **SSD** : Sum of Squared Differences

$$d(q) = \arg \min_d \sum_{p \in F} (I(q + p) - I'(q + de_1 + p))^2$$

- **CSSD** : Centered Sum of Squared Differences

$$d(q) = \arg \min_d \sum_{p \in F} (I(q + p) - \bar{I}_F - I'(q + de_1 + p) + \bar{I}'_F)^2$$

- **NCC** : Normalized Cross-Correlation

$$d(q) = \arg \max_d \frac{\sum_{p \in F} (I(q + p) - \bar{I}_F)(I'(q + de_1 + p) - \bar{I}'_F)}{\sqrt{\sum_{p \in F} (I(q + p) - \bar{I}_F)^2} \sqrt{\sum_{p \in F} (I'(q + de_1 + p) - \bar{I}'_F)^2}}$$

Inconvénients des méthodes locales

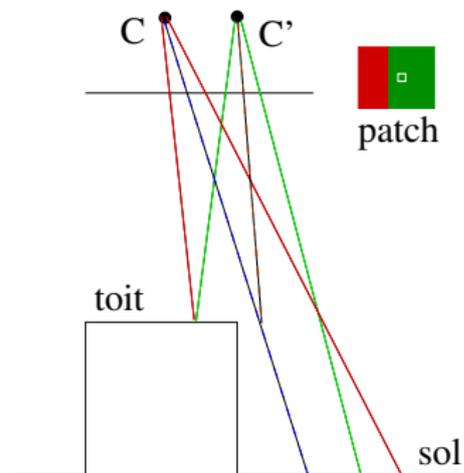
- Hypothèse implicite : tous les points de la fenêtre ont le même mouvement, ie ils sont dans un plan fronto-parallèle.
- Problème d'**ouverture** : le contexte est trop restreint par endroits, manque d'information.
- Problème d'**adhérence** : les discontinuités d'intensité influencent fortement la disparité estimée et si elles correspondent à une discontinuité de profondeur, on a tendance à dilater les objets de premier plan.



- **O** : problème d'ouverture
- **A** : problème d'adhérence

Adhérence

- La fenêtre contient des projections de points avec des profondeurs très différentes.
- Le point central est la projection d'un point au sol.
- Si le toit et le sol sont peu texturés, la discontinuité d'intensité influence fortement la corrélation.



- Le déplacement apparent du point central est la disparité correspondant à un point du toit.
- On a donc une dilatation apparente du toit dans la carte de disparités.
- La dilatation est égale au rayon de la fenêtre.

Exemple de block-matching sans filtrage

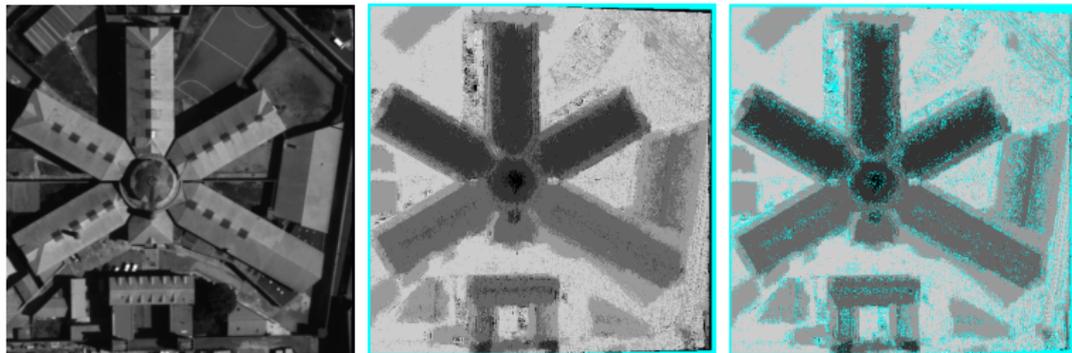


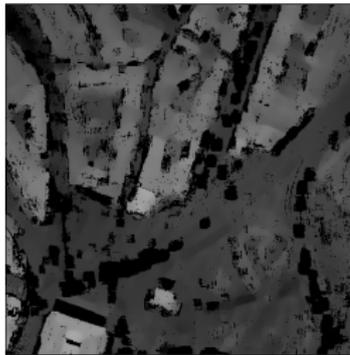
Image originale, carte de disparité, après filtrage des cas d'égalité

Problème de mouvement

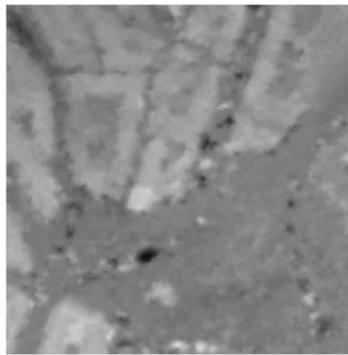


- (a) image gauche
- (b) image droite
- (c) disparité
- (d) masque
- (e) autre méthode

(a)



(b)

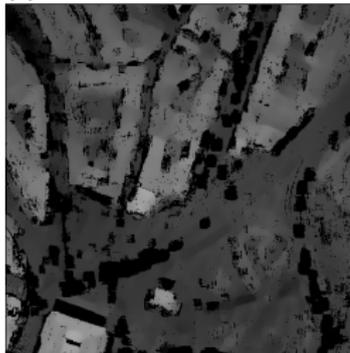


Problème de mouvement

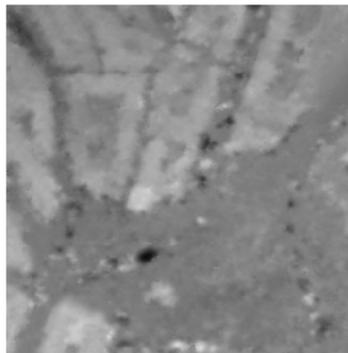


- (a) image gauche
- (b) image droite
- (c) disparité
- (d) masque
- (e) autre méthode

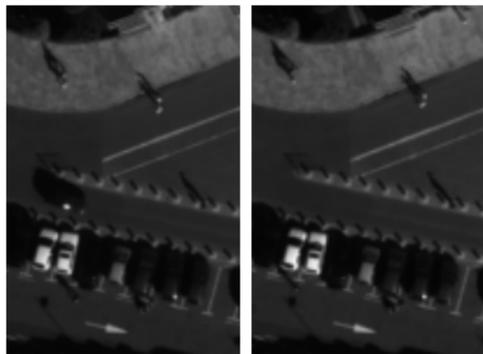
(a)



(b)

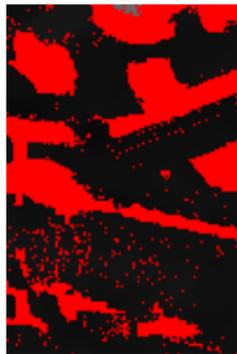


Problème de mouvement

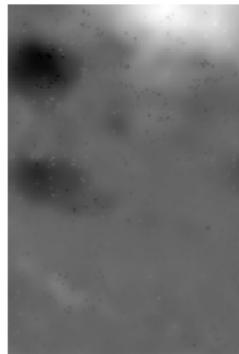


- (a) image gauche
- (b) image droite
- (c) disparité
- (d) masque
- (e) autre méthode

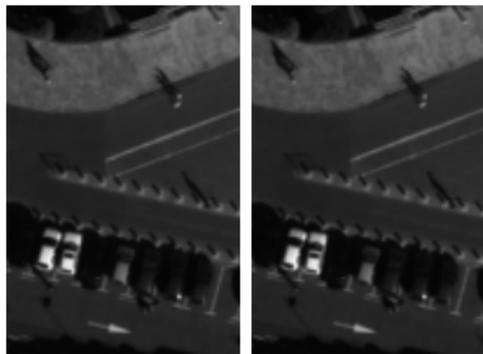
(a)



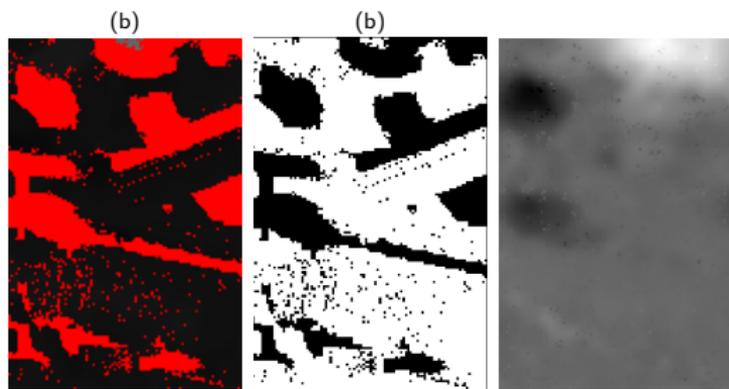
(b)



Problème de mouvement



- (a) image gauche
- (b) image droite
- (c) disparité
- (d) masque
- (e) autre méthode



Modèle *a contrario* pour le Block-matching (AC)

- Nous allons considérer un modèle de patch aléatoire avec un générateur de patch linéaire : $B = c_1 B_1 + \dots + c_n B_n$ avec (B_1, \dots, B_n) famille orthonormale fixe de patches et les c_i des variables aléatoires
- Idée naïve : $B_i = \delta_{i/9, i\%9}$, les c_i sont très dépendants pour avoir un modèle réaliste.
- Mieux : analyse en composantes principales des patches

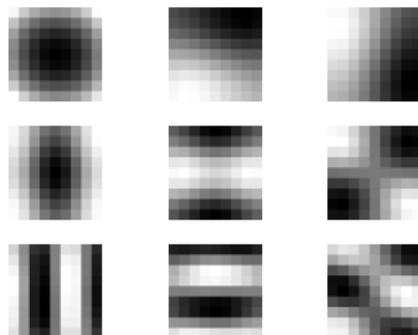


Image et ses premières composantes principales des patches 9×9 .

Modèle *a contrario* pour le block-matching (AC)

Définition : modèle *a contrario*

Le modèle *a contrario* associé à une image de référence I est un vecteur aléatoire $\mathbf{c}(\mathbf{q}) = (\mathbf{c}_1(\mathbf{q}), \dots, \mathbf{c}_N(\mathbf{q}))$ tel que

- $\forall \mathbf{q} \in I$, les composantes $\mathbf{c}_i(\mathbf{q})$, $i = 1, \dots, N$ sont des variables aléatoires indépendantes
- $\forall i$, la loi de $\mathbf{c}_i(\mathbf{q})$ est l'histogramme empirique de $\mathbf{c}_i(\cdot)$ pour l'image I .
- Une **Analyse en Composantes Principales** (ACP) assure que les coefficients $\mathbf{c}_i(\mathbf{q})$ et $\mathbf{c}_j(\mathbf{q})$ sont décorrélés pour $i \neq j \rightarrow$ première approximation d'indépendance
- Réduction de dimension : N premières composantes sont conservées.

Modèle *a contrario* pour le Block-matching (AC)

Définition : Nombre de fausses alarmes

Le nombre de fausses alarmes de l'événement "un bloc aléatoire $B_{q'}$ est aussi similaire à B_q que $B_{q'}$ l'est" se définit

$$NFA(B_q, B_{q'}) = N_{test} \cdot P_{qq'} ;$$

où N_{test} est le nombre de test réalisés et $P_{qq'}$ est la probabilité que $B_{q'}$ soit aussi similaire à B_q que $B_{q'}$ l'est.

Définition : correspondances significatives

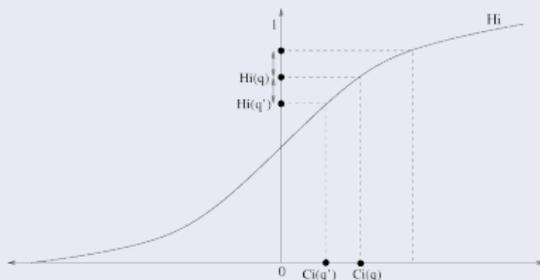
(q, q') est une correspondance ϵ -significative si $NFA(B_q, B_{q'}) \leq \epsilon$

Calcul de probabilités

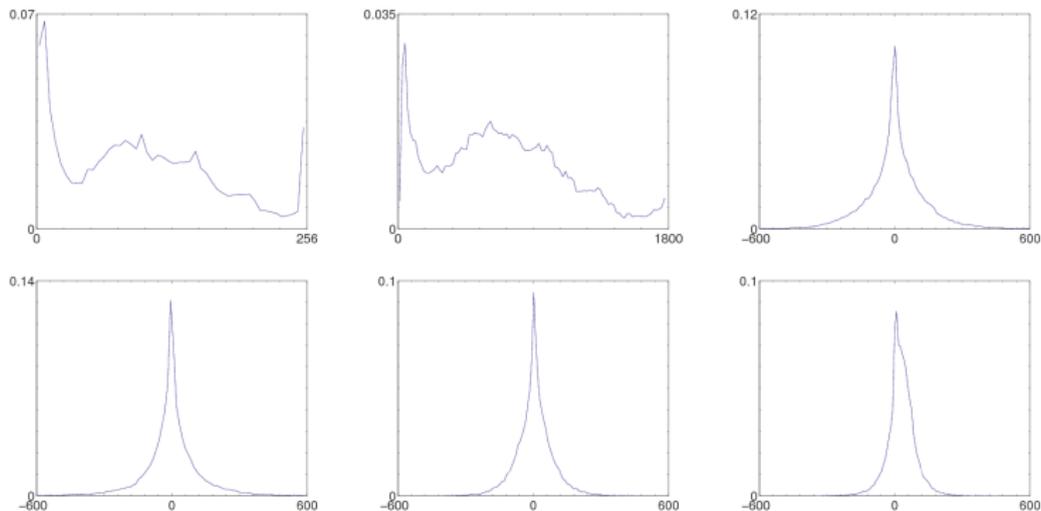
Définition : probabilités empiriques

Soit H_i l'histogramme cumulé de c_i . La probabilité empirique qu'un bloc $B_{\mathbf{q}'}$ soit similaire à $B_{\mathbf{q}}$ pour la i -ème composante est

$$\hat{p}_{\mathbf{q}\mathbf{q}'}^i = \begin{cases} H_i(\mathbf{q}') & \text{si } H_i(\mathbf{q}) < |H_i(\mathbf{q}) - H_i(\mathbf{q}')| \\ 1 - H_i(\mathbf{q}') & \text{si } 1 - H_i(\mathbf{q}) < |H_i(\mathbf{q}) - H_i(\mathbf{q}')| \\ 2 \cdot |H_i(\mathbf{q}) - H_i(\mathbf{q}')| & \text{autrement} \end{cases}$$

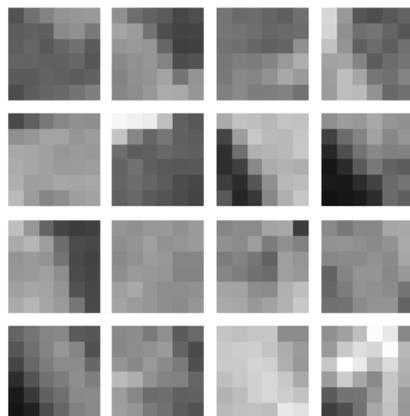


Histogrammes de quelques composantes

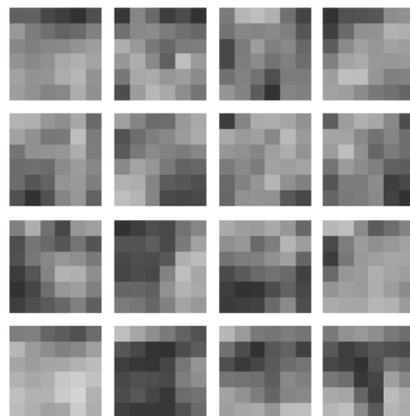


Histogramme de l'intensité de l'image et histogrammes des 5 premières composantes principales

Validité des probabilités empiriques



(a)



(b)

- (a) Patches de l'image de référence, choisis au hasard.
- (b) Patches simulés suivant les lois empiriques

Calcul de la probabilité

- On veut que les k premières composantes aient une faible probabilité :

$$\hat{p}_{qq'}(k) = \max_{i=1\dots k} \hat{p}^i_{qq'}. \quad (1)$$

- On quantifie par une fonction π envoyant vers $\frac{1}{2^{Q-1}}, \dots, \frac{1}{2}, 1$:

$$\pi \left(\max_{i=1\dots k} \hat{p}^i_{qq'} \right) = \pi_j \quad (2)$$

- Si on les considère indépendantes, on obtient une probabilité

$$\pi_j^k \quad (3)$$

- Comme on ne sait pas combien de composantes considérer, on essaie plusieurs k et on retient le meilleur :

$$P_{\mathbf{q}\mathbf{q}'} = \min_{k=k_{\min}\dots k_{\max}} \pi \left(\max_{i=1\dots k} \hat{p}^i_{qq'} \right)^k \quad (4)$$

Nombre de tests

$$N_{test} = \#I \cdot \#disp \cdot Q \cdot (k_{max} - k_{min} + 1) \quad (5)$$

- $\#I$: nombre de pixels dans l'image I
- $\#disp$: nombre de disparités testées
- Q : nombre de niveaux de quantification des probabilités
- k_{min}, k_{max} : nombre min et max de composantes testées

Modèle *a contrario* pour le block-matching (AC)

Théorème 1

Soit

$$\mathbf{1}_{qq'} = \begin{cases} 1, & \text{if } NFA(B_q, \mathbf{B}_{q'}) \leq \epsilon; \\ 0, & \text{if } NFA(B_q, \mathbf{B}_{q'}) > \epsilon. \end{cases}$$

et $\Gamma = \sum_{q,q'} \mathbf{1}_{B_q, \mathbf{B}_{q'}}$ la variable aléatoire qui représente le nombre d'occurrences d'une correspondance ϵ -significative entre un bloc déterministe de la première image et un bloc aléatoire de la deuxième. Alors

$$\mathbb{E}[\Gamma] \leq \epsilon$$

Solution à l'effet stroboscopique

- Erreurs dues aux structures répétées dans les images.
- Solution : le **seuil d'autosimilarité** (adaptation du seuil SIFT)

AC+SA

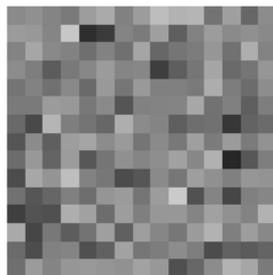
Une correspondance $(\mathbf{q}, \mathbf{q}')$ est retenue si

- $(\mathbf{q}, \mathbf{q}')$ est une correspondance significative ;
- $d(B_{\mathbf{q}}, B_{\mathbf{q}'}) < \lambda \min\{d(B_{\mathbf{q}}, B_{\mathbf{r}}) \mid \mathbf{r} \in I \cap S(\mathbf{q})\}$,

où $S(\mathbf{q}) = [q_1 - R, q_1 + R] \setminus \{q_1 - 1, q_1, q_1 + 1\}$ avec $\mathbf{q} = (q_1, q_2)$ et R la distance de recherche.

En pratique, on prend $\lambda = 1$.

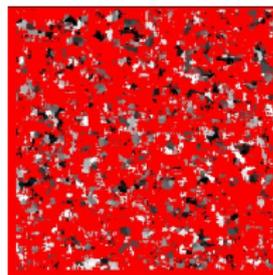
Besoin de AC + SA



(a) Image



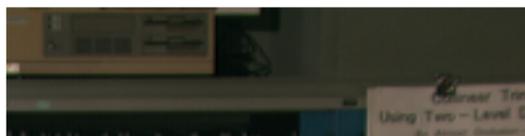
(b) AC



(c) SA

- (a) Image de bruit
- (b) Aucune correspondance validée par le test *a contrario*
- (c) Beaucoup de fausses correspondances acceptées par SA

Besoin de AC + SA



image



SA



AC

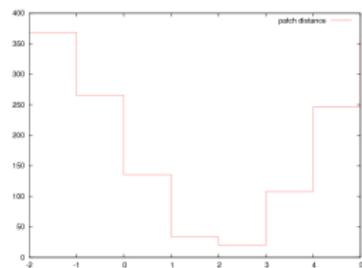
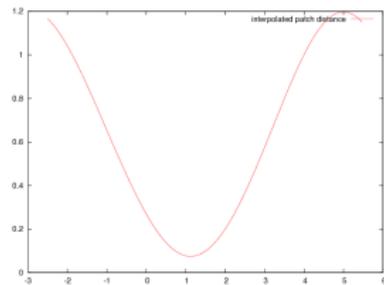
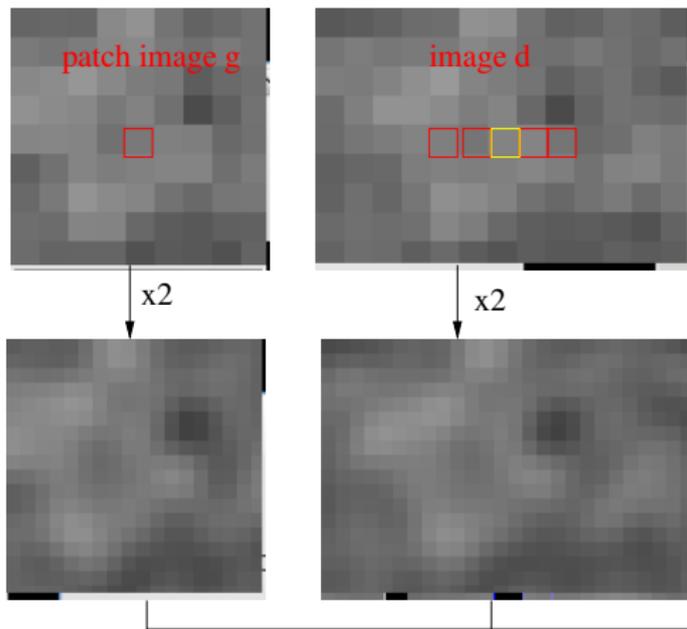


AC+SA

Disparité sous-pixellique

- Recherche sous-pixellique (1/64 pixel) du maximum de corrélation
- Très coûteux en temps et mémoire de faire un zoom $\times 64$ des images
- Zoom $\times 2$ des images et corrélations entières : la fonction corrélation $e : \mu \rightarrow \langle u_g(\cdot, \cdot + \mu), u_d \rangle$ est à Shannon si u_g et u_d le sont
- Zoom $\times 32$ de $e(\mu)$ pour trouver son maximum

Disparité sous-pixellique



Justification

- Discrete Fourier Transform (DFT) :

$$u(x, y) = \sum_{k, l \in \mathbb{Z}} \tilde{u}_{k, l} e^{\frac{2i\pi(kx+ly)}{a}}, \quad (6)$$

- The DFT is an isometry :

$$\int_{[0, a]^2} u(\mathbf{x}) \bar{v}(\mathbf{x}) d\mathbf{x} = a^2 \sum_{k, l \in \mathbb{Z}} \tilde{u}_{k, l} \overline{\tilde{v}_{k, l}}. \quad (7)$$

- If u and v are N -degree trigonometric polynomials :

$$\int_{[0, a]^2} u(\mathbf{x}) \bar{v}(\mathbf{x}) = a^2 \sum_{k, l = -N}^{N-1} \tilde{u}_{k, l} \overline{\tilde{v}_{k, l}} = \frac{a^2}{4N^2} \sum_{\mathbf{m} \in \mathbb{Z}_a^{\frac{1}{2}}} u(\mathbf{m}) \overline{v(\mathbf{m})}. \quad (8)$$

Justification

Recherche de la correspondance :

$$e_{\mathbf{x}_0}(\mu) := \int_{[0,a]^2} \varphi(\mathbf{x} - \mathbf{x}_0) (u_1(\mathbf{x}) - u_2(\mathbf{x} + (\mu, 0)))^2 d\mathbf{x}. \quad (9)$$

Proposition : Equality of the discrete and the continuous quadratic distance

Let $u_1(\mathbf{x})$ and $u_2(\mathbf{x})$ be two N -degree trigonometric polynomials on $[0, a]^2$ and let $\varphi(\mathbf{x})$ be a window function which we assume to be a $2N$ -degree trigonometric polynomial. Then

$$e_{\mathbf{x}_0}(\mu) = e_{\mathbf{x}_0}^d(\mu), \quad \text{where} \quad (10)$$

$$e_{\mathbf{x}_0}^d(\mu) := \frac{a^2}{4N^2} \sum_{\mathbf{m} \in \mathbb{Z}_2^{\frac{1}{2}}} \varphi(\mathbf{m} - \mathbf{x}_0) (u_1(\mathbf{m}) - u_2(\mathbf{m} + (\mu, 0)))^2. \quad (11)$$

Justification

Proposition : Sub-pixel correlation requires $\times 2$ zoom

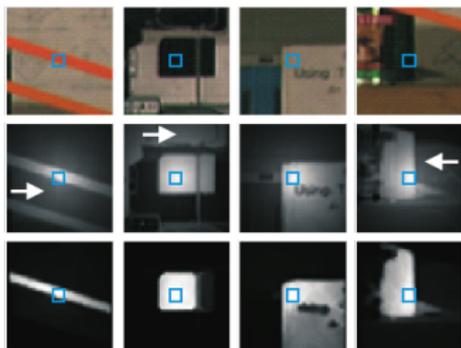
Let $u_1(\mathbf{x})$ and $u_2(\mathbf{x})$ be two N -degree trigonometric polynomials. Then the quadratic distance $e_{\mathbf{x}_0}^d(\mu)$ is well-sampled provided it has at least $2N$ successive samples. Thus the computation of $e_{\mathbf{x}_0}^d(\mu)$ at half samples $\mu \in \frac{a\mathbb{Z}}{2}$ (via zero-padding) allows the exact reconstruction of $e_{\mathbf{x}_0}^d(\mu)$ for any real μ by DFT interpolation.

Conclusion

- Le calcul de cartes de disparités est un thème de recherche très actif : voir le benchmark Middlebury, avec plus de 100 algorithmes. <http://vision.middlebury.edu/stereo/>
- Une méthode de rectification sous IPOL : http://www.ipol.im/pub/algo/m_quasi_euclidean_epipolar_rectification/
- La chaîne stéréo telle que présentée dans ce cours : http://dev.ipol.im/~monasse/ipol_demo/bms_binocular_stereo_pipeline/

Projet 1 : algorithme de stéréo avec fenêtres adaptatives pour IPOL

- A. Hosni, M. Bleyer, M. Gelautz, C. Rhemann *Local stereo matching using geodesic support weights*. ICIP, 2009
- Définition de voisinages adaptatifs par un poids (entre 0 et 1) attribué à chaque pixel d'une fenêtre suffisamment grande : ce poids est une fonction décroissante de la géodésique en variation totale pour joindre le point central.



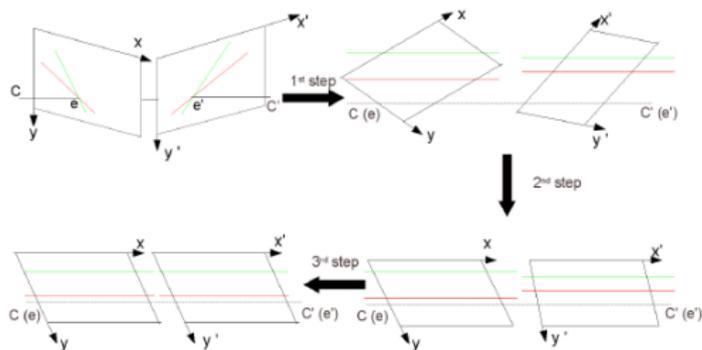
Haut : patches de l'image.

Milieu : poids de Yoon et Kweon,
CVPR 2005.

Bas : poids géodésiques

Projet 2 : Rectification par itération de 3 rotations successives

- P. Monasse, J.-M. Morel et Z. Tang. *Three-step image rectification*. BMVC 2010.
- Modèle simplifié de calibration avec uniquement une focale inconnue, à la manière de Fusiello-Irsara
http://www.ipol.im/pub/algo/m_quasi_euclidean_epipolar_rectification/



Minimisation d'une erreur non linéaire de l'unique inconnue f : possibilité d'optimisation globale.

Projet 3 : Rectification de Hartley

- R.I. Hartley. *Theory and practice of projective rectification*. IJCV, 1999
- Calcul de F à partir de correspondances
- Tourne les images et envoie les épipoles à l'infini dans la direction x
- Transformation affine $x' = ax + by + c$ pour minimiser les disparités

Projet 4 : Rectification de Loop et Zhang

- C. Loop et Z. Zhang. *Computing rectifying homographies for stereo vision*. CVPR, 1999
- 3 étapes : projectif, similitude et cisaillement (shear), chacun minimisant la distorsion
- **Projectif** : envoie e à l'infini et garde un point z à l'infini fixe. Extraction de racine d'un polynôme de degré 7 pour trouver z .
- **Similitude** : envoie e en $(0 \ 0 \ 1)^T$
- **Cisaillement** : conserve l'orthogonalité des lignes médianes.