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

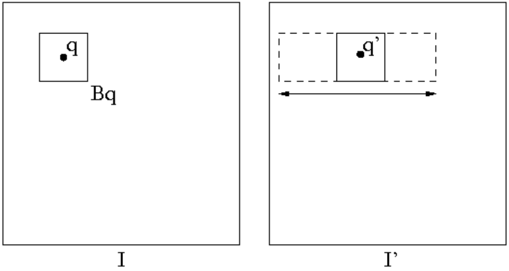
Neus Sabater, Pascal Monasse, Jean-Michel Morel

19 novembre 2010

Plan

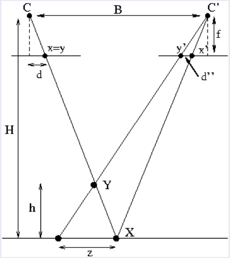
Stéréoscopie et mise en correspondance des images

- 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.



Stéréoscopie et mise en correspondance des images

Principe fondamental de la vision stéréoscopique



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

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. $\longrightarrow$ 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.

Exemple de block-matching sans filtrage

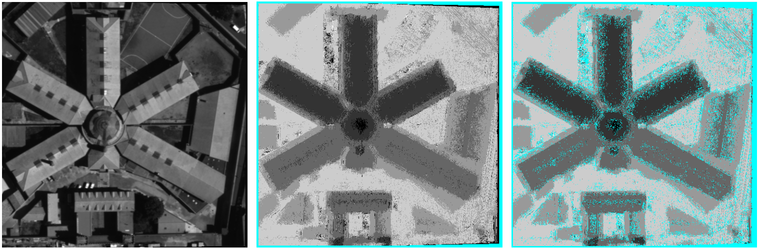
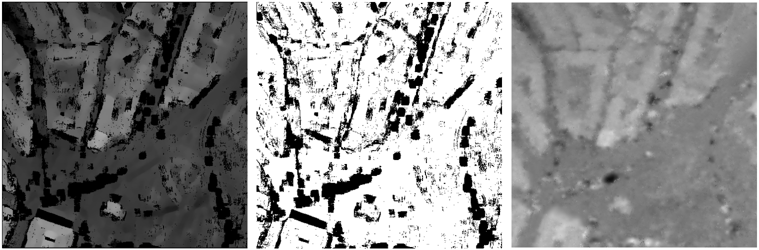


Image originale, carte de disparité, après filtrage des cas d'égalité

Problème de mouvement

(a) (b)



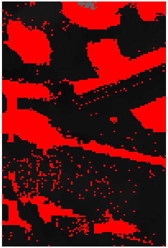
- (a) image gauche
- (b) image droite
- (c) disparité
- (d) masque
- (e) autre méthode

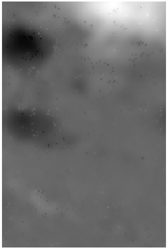
Problème de mouvement


(a)

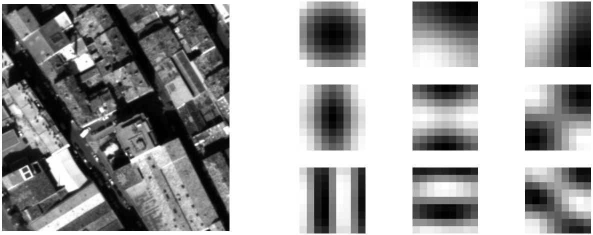

(b)

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





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



Une image et les 9 premières composantes principales des patchs 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 $\mathbf{c}(\mathbf{q}) = (\mathbf{c}_1(\mathbf{q}), \ldots, \mathbf{c}_N(\mathbf{q}))$ tel que

- $\forall \mathbf{q} \in I$, les composantes $\mathbf{c}_i(\mathbf{q})$, $i = 1, \ldots, N$ sont des variables aléatoires indépendantes
- $\forall i$, la loi de $\mathbf{c}_i(\mathbf{q})$ est l'histogramme empirique de $c_i(\cdot)$ 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 $i \neq j \longrightarrow$ 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 $\mathbf{B_{q'}}$ est aussi similaire à $B_{\mathbf{q}}$ que $B_{\mathbf{q'}}$ l'est" se définit

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

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

Définition : correspondances significatives

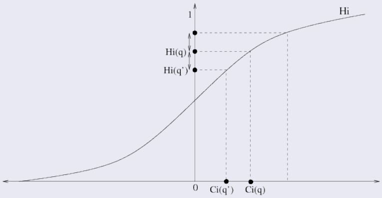
$(\mathbf{q}, \mathbf{q}')$ est une correspondance ϵ-significative si $NFA(B_{\mathbf{q}}, B_{\mathbf{q'}}) \leqslant \epsilon$

Calcul de probabilités

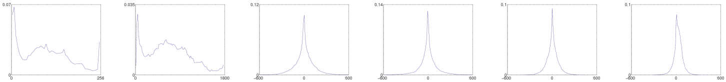
Définition : probabilités empiriques

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 composante i-ème est

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

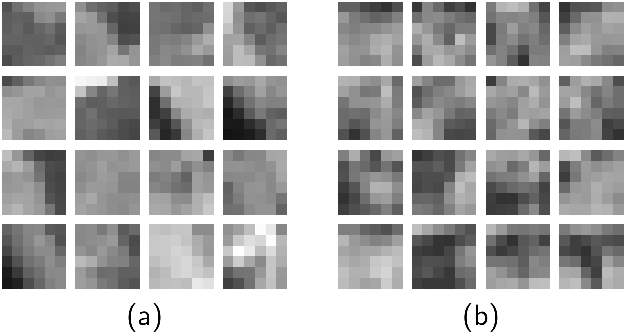


Histogrammes de quelques composantes



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

Validité des probabilités empiriques



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

Calcul de la probabilité

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

$$\widehat{p}_{B_\mathbf{q}\,B'}(k) = \max_{i=1\ldots k} \widehat{p^i}_{B_\mathbf{q}\,B'}. \tag{1}$$

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

$$\pi\left(\max_{i=1\ldots k} \widehat{p^i}_{B_\mathbf{q}\,B'}\right) = \pi_j \tag{2}$$

- Si on les considère indépendantes, on obtient une probabilité

$$\pi_j^k \tag{3}$$

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

$$P_{\mathbf{q}\,\mathbf{q}'} = \min_{k=k_{\min}\ldots k_{\max}} \pi\left(\max_{i=1\ldots k} \widehat{p^i}_{B_\mathbf{q}\,B'}\right)^k \tag{4}$$

Nombre de tests

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

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

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

Théorème 1

Soit

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

et $\Gamma = \Sigma_{\mathbf{q},\mathbf{q}'} \mathbf{1}_{B_{\mathbf{q}},\mathbf{B}_{\mathbf{q}'}}$ 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

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

Solution à l'effet stroboscopique

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

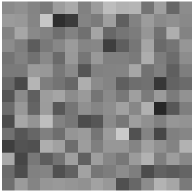
AC+SA

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

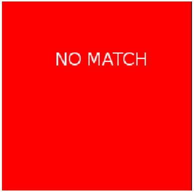
- $(\mathbf{q}, \mathbf{q}')$ est une correspondance significative ;
- $d(B_{\mathbf{q}}, B_{\mathbf{q}'}) < 0.6 \ \min\{d(B_{\mathbf{q}}, B_{\mathbf{r}}) |\ r \in I \cap S(\mathbf{q})\}$,

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.

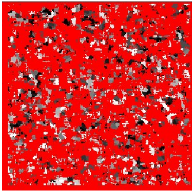
Besoin de AC + SA



(a) Image



(b) AC

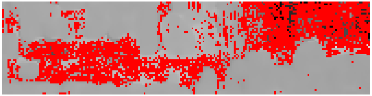


(c) SA

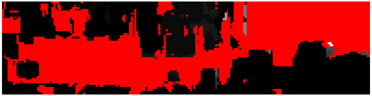
- (a) Image de bruit
- (b) Aucune correspondance validée par le test *a contrario*
- (c) Beaucoup de fausses correspondances acceptées par SA

Besoin de AC + SA

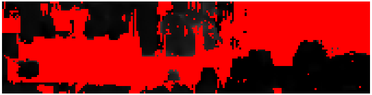
image



AC



SA

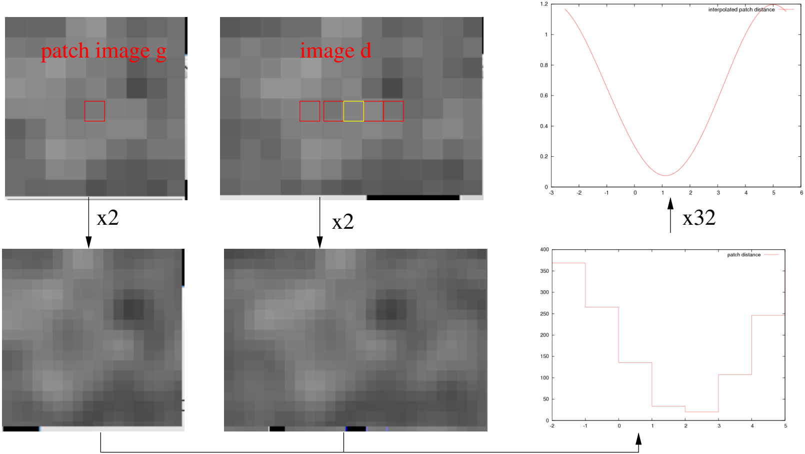


AC+SA

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 $e : \mu \rightarrow \langle u_g(.,.+\mu), u_d \rangle$ 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) :

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

- The DFT is an isometry :

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

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

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

Justification

Recherche de la correspondance :

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

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

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

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

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 $e^d_{\mathbf{x}_0}(\mu)$ is well-sampled provided it has at least $2N$ successive samples. Thus the computation of $e^d_{\mathbf{x}_0}(\mu)$ at half samples $\mu \in \frac{a\mathbb{Z}}{2}$ (via zero-padding) allows the exact reconstruction of $e^d_{\mathbf{x}_0}(\mu)$ for any real μ by DFT interpolation.

Projet 1 : implémentation de divers filtres dans IPOL

- D. Scharstein and R. Szeliski. *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms.* International Journal of Computer Vision, 47(1/2/3) :7-42, April-June 2002.
- compatibilité gauche→droite et droite→gauche
- le min-filter : la fenêtre n'est pas forcément centrée
- autres filtres...

Projet 2 : algorithme RAFA

- Enlève le problème d'adhérence (fattening) en-dehors des discontinuités.
- Solution : SSD avec chaque pixel pondéré par le carré de la dérivée suivant x