

# Estimation itérative des profondeurs et du mouvement de caméra

Georges Koepfler

Université Paris Descartes

MVA 2010-2011

## Plan :

- 1 Estimation des Profondeurs
- 2 Profondeurs et mouvement

Avec la longueur focale on a :

$$x = f_c \frac{X}{Z} = f_c \frac{a_1 x' + b_1 y' + f_c c_1 + f_c t_1 / Z'(x', y')}{a_3 x' + b_3 y' + f_c c_3 + f_c t_3 / Z'(x', y')},$$

$$y = f_c \frac{Y}{Z} = f_c \frac{a_2 x' + b_2 y' + f_c c_2 + f_c t_2 / Z'(x', y')}{a_3 x' + b_3 y' + f_c c_3 + f_c t_3 / Z'(x', y')},$$

et

$$g(x', y') = f(x, y) = f(\varphi(x', y'))$$

Si on connaît le déplacement  $D = (R, t)$ , alors en comparant  $g$  à  $D(f)$ , on peut estimer une **profondeur moyenne** dès que  $t \neq 0$ .

## Profondeurs par BP

- L'observation sur l'ensemble des pixels  $K$  est

$$y = \{f, g, \theta, \alpha, \beta, t_1, t_2, t_3\} = \{y_s, s \in K\}$$

$y$  est considérée comme la réalisation d'un champ aléatoire  $Y$ .

- On souhaite estimer  $\mathbf{x} = \{x_s, s \in K\}$  qui est une réalisation du champ de profondeurs  $X$  vue de l'image  $f$  en maximisant la probabilité *a posteriori*

$$P(\mathbf{x}) = P(X = \mathbf{x} | Y = y)$$

- L'estimation bayésienne de  $\mathbf{x}$  est basée sur

$$P(\mathbf{x}) = P(X = \mathbf{x} | Y = y) \propto P(Y = y | X = \mathbf{x}) P(X = \mathbf{x}).$$

On a

$$P(Y = y|X = x) = \prod_{s \in K} \psi(x_s)$$

où  $K$  est l'ensemble de pixels,  $x_s$  la profondeur relative au pixel  $s$ .

La *vraisemblance locale* ou *attache au données*,  $\psi(x_s)$  est définie comme

$$\psi(x_s) = (1 - e_2)e^{-\frac{F(s)}{\sigma^2}} + e_2$$

où  $F(s) = |f(s) - g(s')|$  et  $s' = (x', y')$  est le pixel associé à  $s = (x, y)$  grâce à l'estimation du déplacement  $\{\theta, \alpha, \beta, t_1, t_2, t_3\}$  si la profondeur du point 3D projeté en  $s$  est  $x_s$ .

## Attache aux données (suite)

Pour  $s$  et  $s'$  on a :

$$g(x', y') = f \left( f_c \frac{a_1 x' + b_1 y' + f_c c_1 + f_c \frac{t_1}{x_s}}{a_3 x' + b_3 y' + f_c c_3 + f_c \frac{t_3}{x_s}}, f_c \frac{a_2 x' + b_2 y' + f_c c_2 + f_c \frac{t_2}{x_s}}{a_3 x' + b_3 y' + f_c c_3 + f_c \frac{t_3}{x_s}} \right)$$

On suppose que la profondeur moyenne de la scène  $Z_0 = 1$ , alors

Si  $x_s < 1$ , le pixel  $s$  est en avant du plan de profondeur  $Z_0$ .

Si  $x_s > 1$ , le pixel  $s$  en arrière du plan de profondeur  $Z_0$ .

Noter que la longueur focale peut être déduite de la taille de l'image  $L$  et de l'angle de vue  $a$  :

$$f_c = \frac{L}{2 \tan(a/2)}.$$

La connaissance *a priori* sur  $X$  s'écrit

$$P(X = \mathbf{x}) = \prod_{(s,t) \in E} \psi_{st}(\mathbf{x}_s, \mathbf{x}_t).$$

où  $E$  est l'ensemble des couples de pixels  $(s, t)$  4-voisins.

La fonction compatibilité  $\psi_{st}(\mathbf{x}_s, \mathbf{x}_t)$  est donnée par

$$\psi_{st}(\mathbf{x}_s, \mathbf{x}_t) = (1 - e_1) e^{-\frac{|\mathbf{x}_s - \mathbf{x}_t|}{\sigma_1}} + e_1$$

Elle impose de la régularité sur le champ de profondeurs.

## Exemples d'estimation de la profondeur

On va comparer

- (1) Estimation de  $D$  avec  $\mathbf{x}_s = 1$  ;
- (2) Estimation de  $\tilde{D}$  avec  $\mathbf{x}_s \in \Lambda$ , déterminés par BP

Ensemble fini de profondeurs relatives :

$$\Lambda = \{0.45; 0.5; 0.55; 0.6; 0.65; 0.7; 0.75; 0.8; 0.9; 1; 1.1; 1.2; 1.4; 1.6; 1.8; 2\}$$

Paramètres pour BP :  $e_1 = 0.01$ ,  $\sigma_1 = 0.3$ ,  $e_2 = 0.05$  and  $\sigma_2 = 20$ .

On améliore le recalage, mais la carte des profondeurs n'est pas satisfaisante.

# Exemple 1



$$\|D(f) - g\|_1 = 20.60$$

100 itérations BP

$$\|\tilde{D}(f) - g\|_1 = 9.92$$

# Exemple 2



$$\|D(f) - g\|_1 = 15.38$$

100 itérations

$$\|\tilde{D}(f) - g\|_1 = 7.07$$

- Partition de l'ensemble des profondeurs relatives  $\Lambda = I_1 \dots \cup I_H$  ;
- Soit  $\mathbf{x}$  le champ de profondeurs :  
on obtient une partition de  $K$  par la relation  $\mathbf{x}_{(x,y)} \in I_h$  ;
- La profondeur moyenne des pixels de la classe  $h$  est  
$$\hat{X}_h = \text{moyenne de } I_h ;$$
- La fraction de l'image de profondeur moyenne  $\hat{X}_h$  est  
$$p_h = |\{\mathbf{x} \in I_h\}|/|K| ;$$
- Le flot optique est calculé par Motion2D pour tous les pixels de  $K$  tels que  $\mathbf{x}_{(x,y)} \in I_h$ , en utilisant  $\hat{X}_h$ .  
On obtient les paramètres  $(c_1^h, c_2^h, a_1^h, a_2^h, q_1^h, q_2^h)$  ;
- À partir de  $\{(c_1^h, c_2^h, a_1^h, a_2^h, q_1^h, q_2^h), 1 \leq h \leq H\}$ , on calcule  $(\theta, \alpha, \beta, A, B, C)$  en utilisant les poids  $p_h$  pour pondérer.

## Algorithme itératif

### 1 Initialisation

- estimer le déplacement  $D_1$  entre  $f$  et  $g$  ;
- estimer le champ de profondeurs  $\mathbf{x}_1$  par BP grâce à  $f$ ,  $g$  et  $D_1$  ;
- on note  $\tilde{\varphi}$  la transformation associé à  $D_1$  et  $\mathbf{x}_1$ .

### 2 Pour $i = 2, \dots, N$

- pour  $h = 1, \dots, H$ 
  - estimer le mouvement 2D sur  $K \cap \{\mathbf{x}_{i-1} \in I_h\}$  ;
  - calculer les poids  $p_h = |\{\mathbf{x}_{i-1} \in I_h\}|/|K|$  ;
- déterminer  $D_i$  à partir des  $H$  mouvements 2D, en utilisant les poids  $p_h$  ;
- estimer le champ de profondeurs  $\mathbf{x}_i$  par BP grâce à  $f$ ,  $g$  et  $D_i$  ;
- on note  $\varphi_i$  la transformation associé à  $D_i$  et  $\mathbf{x}_i$  ;
- si  $\|f \circ \varphi_i - g\|_1 < \|f \circ \tilde{\varphi} - g\|_1$ , on pose  $\tilde{\varphi} = \varphi_i$ .

Partition de l'ensemble des profondeurs relatives  $\Lambda = I_1 \cup I_2 \cup I_3$  :

$$I_1 = \{0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75\},$$

$$I_2 = \{0.8, 0.9, 1, 1.1, 1.2\},$$

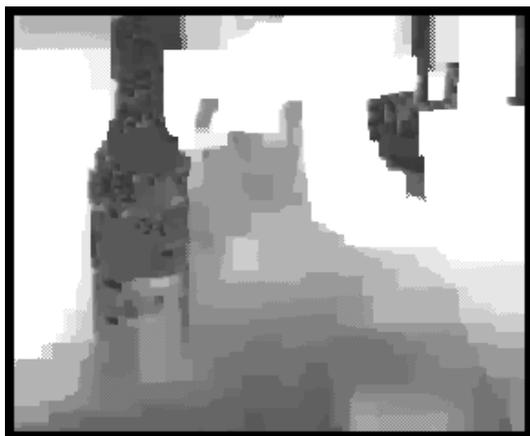
$$I_3 = \{1.4, 1.6, 1.8, 2\}.$$

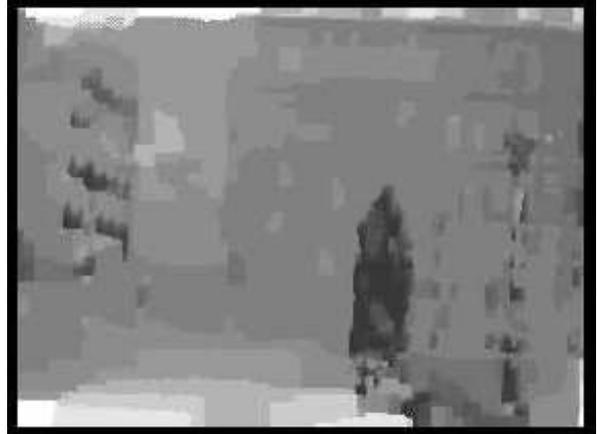
Profondeurs moyennes :

$$\hat{X}_1 = 0.6, \hat{X}_2 = 1 \text{ and } \hat{X}_3 = 1.7$$

$N = 15$  itérations de l'algorithme et l'algorithme de BP à chaque fois 100 itérations.

## Exemple 1





## Références

Ces notes sont basées sur les documents suivants :

- C. Jonchery, PhD dissertation, *Estimation d'un mouvement de caméra et problèmes connexes*, ENS Cachan, November 2006.
- F. Dibos, C. Jonchery, G. Koepfler, *Camera motion estimation through planar deformation determination*, in Journal of Mathematical Imaging and Vision, vol 32, 1, p. 73-87, September 2008.
- F. Dibos, C. Jonchery, G. Koepfler, *Iterative Camera Motion and Depth Estimation in a Video Sequence*, in Jiang, X. and Petkov, N. (ed.) CAIP 2009. LNCS, vol. 5702, pp. 1028–1035, Springer, Heidelberg, September 2009.