

# flot optique

enric meinhardt-llopis  
javier sánchez  
gabriele facciolo

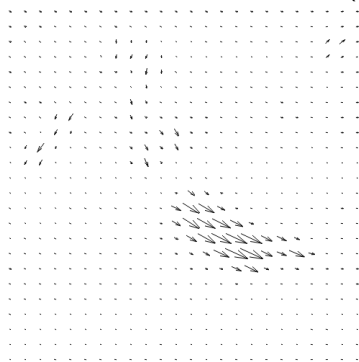
11 – 4 – 2011

# Le problème du flot optique

Entrée : vidéo  $I_i(\mathbf{x})$



Sortie : champs vectoriels  $\mathbf{u}_i(\mathbf{x})$



Critère :  $I_{t+1}(\mathbf{x} + \mathbf{u}_t(\mathbf{x})) = I_t(\mathbf{x})$

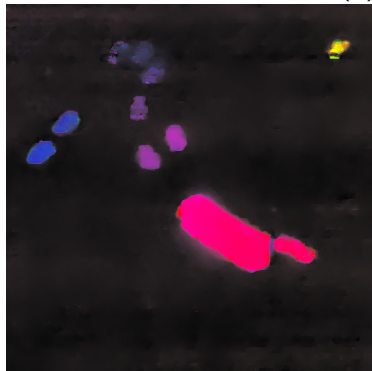
Critère linearisé :  $\nabla I \cdot \mathbf{u} + \frac{\partial}{\partial t} I = 0$

# Le problème du flot optique

Entrée : vidéo  $I_i(\mathbf{x})$



Sortie : champs vectoriels  $\mathbf{u}_i(\mathbf{x})$



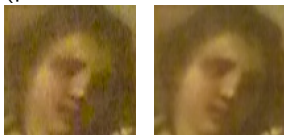
Critère :  $I_{t+1}(\mathbf{x} + \mathbf{u}_t(\mathbf{x})) = I_t(\mathbf{x})$

Critère linearisé :  $\nabla I \cdot \mathbf{u} + \frac{\partial}{\partial t} I = 0$

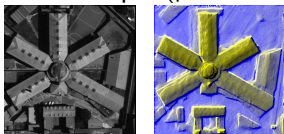
# Applications du flot optique

Chaque application a des réqueriments très différents

- ▶ Suivi d'objets en vidéos (temps réel, localement précis)
- ▶ Stabilisation de l'image (temps réel, globalement précis)
- ▶ Compression de vidéo (bonne compression)
- ▶ Recalage non paramétrique, débruitage, super-résolution  
(précision maximale autour des **discontinuités de l'image**)



- ▶ Stéréoscopie (précision maximale autour des **discontinuités du flot**)





# Méthodes analysées

Seulement des méthodes variationnelles

Trouver le champ de vecteurs qui minimise les énergies suivantes :

Horn-Schunck : 
$$E(u) = \int |\nabla I \cdot u + I_t|^2 + \alpha^2 \int |\nabla u|^2$$

TV-L1 : 
$$E(u) = \int |\nabla I \cdot u + I_t| + \lambda \int |\nabla u|$$

Lucas-Kanade : 
$$E(u(x)) = \int |\nabla I(x+y) \cdot u(x) + I_t(x+y)|^2 g_\sigma(y) dy$$

Chaque méthode a un paramètre de régularisation.

# Nécessité de la multi-échelle

Ces méthodes ne sont adaptés qu'aux petits déplacements.

Mais tout déplacement est sous-pixelien à une échelle suffisamment petite



taille :  $1920 \times 1080$

déplacement : 100 pixels



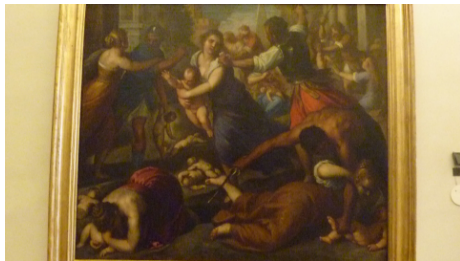
taille :  $15 \times 8$

déplacement : sous-pixelien

# Nécessité de la multi-échelle

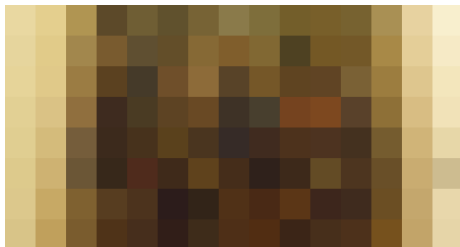
Ces méthodes ne sont adaptés qu'aux petits déplacements.

Mais tout déplacement est sous-pixelien à une échelle suffisamment petite



taille :  $1920 \times 1080$

déplacement : 100 pixels



taille :  $15 \times 8$

déplacement : sous-pixelien

# Méthode de Horn-Schunck

Fonctionnel :  $E(u) = \int |\nabla I \cdot u + I_t|^2 + \alpha^2 \int |\nabla u|^2$

Algorithme : système linéaire  $\implies$  Gauss-Seidel

Paramètres :	interpretation	min	défaut	max
$\alpha$	régularisation	0	10	100
$\eta$	facteur d'échelle	1.1	1.6	3
$N_{\text{first}}$	première échelle	1	6	$1 + \log_{\eta}(\text{taille})$
$N_{\text{last}}$	dernière échelle	1	1	
$\epsilon$	condition d'arrêt	0	0.001	1
$M$	maximum d'itérations	1	500	$\infty$

## Références :

[B.K.P. Horn and B.G. Schunck, Determining Optical Flow, MIT AI MEMO 1980]

[E. Meinhardt and J. Sánchez, Horn-Schunck Optical Flow, IPOL 2012]

# Méthode TV-L1

Attache aux données L1 avec régularisation par variation totale

Fonctionnel :  $E(u) = \int |\nabla I \cdot u + I_t| + \lambda \int |\nabla u|$

Algorithme : séparation de variables, puis méthode de Chambolle  
et seuillage :  $E(u, v) = \int |\nabla I \cdot u + I_t| + \frac{1}{2\theta} |u - v|^2 + \lambda \int |\nabla v|$

Paramètres :	interpretation	min	défaut	max
$\lambda$	régularisation	0	6	100
$\eta$	facteur d'échelle	1.1	1.6	3
$N_{\text{first}}$	première échelle	1	6	$1 + \log_{\eta}(\text{taille})$
$N_{\text{last}}$	dernière échelle	1	1	
$\epsilon, \tau, \theta, \dots$	paramètres internes			$N_{\text{first}}$

## Références :

[A.Chambolle, An algorithm for total variation minimization and applications, JMIV 2004]

[C.Zach, T.Pock and H.Bischof, A duality based approach for realtime TV-L1 Optical Flow, Pat.Rec. 2007]

[J.Sánchez, E.Meinhardt and G.Facciolo, TV-L1 Optical Flow Estimation, IPOL 2012]

# Méthode Lucas-Kanade

Estimation d'un modèle localement constant

Fonctionnel :

$$E(u(x)) = \int |\nabla I(x + y) \cdot u(x) + I_t(x + y)|^2 g_\sigma(y) dy$$

Algorithme : moindres carrées de dimension 2 pour chaque pixel.

Paramètres :	interpretation	min	défaut	max
$\sigma$	amplitude de la fenêtre	1	3	$\infty$
$\eta$	facteur d'échelle	1.1	1.6	3
$N_{\text{first}}$	première échelle	1	6	$1 + \log_\eta(\text{taille})$
$N_{\text{last}}$	dernière échelle	1	1	$N_{\text{first}}$
$k$	taille de la fenêtre	3	5	taille de l'image

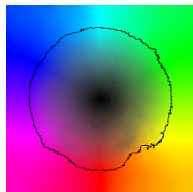
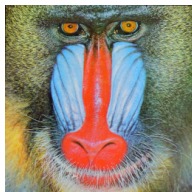
Références :

[B.D. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision, 1981]

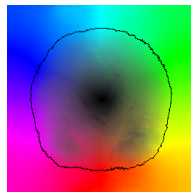
[J.Y.Bouguet, Pyramidal implementation of the affine lucas kanade feature tracker, 2001]

# Comparaison des méthodes

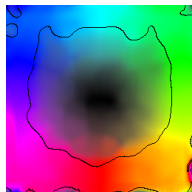
Séquence "Rotating baboon"



HS



TVL1

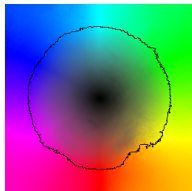
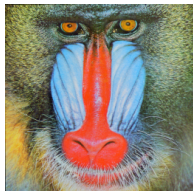


LK

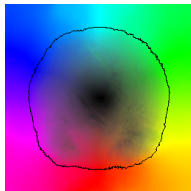
	temps	erreur de réprojection	erreur moyen
V.T.	0.00 s	7.642 niveaux de gris	0.000 pixels
HS	0.24 s	7.530 niveaux de gris	0.069 pixels
TVL1	0.97 s	7.731 niveaux de gris	0.065 pixels
LK	0.14 s	9.158 niveaux de gris	0.164 pixels

# Comparaison des méthodes

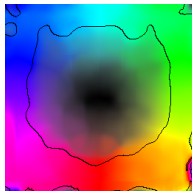
Séquence "Rotating baboon"



HS



TVL1



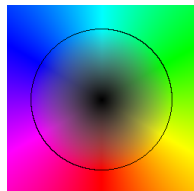
LK

	temps	erreur de réprojection	erreur moyen
V.T.	0.00 s	7.642 niveaux de gris	0.000 pixels
HS	0.24 s	7.530 niveaux de gris	0.069 pixels
TVL1	0.97 s	7.731 niveaux de gris	0.065 pixels
LK	0.14 s	9.158 niveaux de gris	0.164 pixels

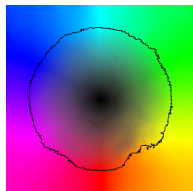


# Comparaison des méthodes

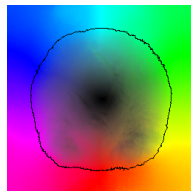
Séquence "Rotating baboon"



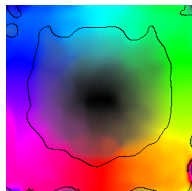
V.T.



HS



TVL1

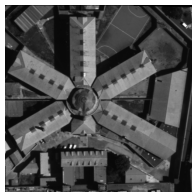


LK

	temps	erreur de réprojection	erreur moyen
V.T.	0.00 s	7.642 niveaux de gris	0.000 pixels
HS	0.24 s	7.530 niveaux de gris	0.069 pixels
TVL1	0.97 s	7.731 niveaux de gris	0.065 pixels
LK	0.14 s	9.158 niveaux de gris	0.164 pixels

# Comparaison des méthodes

Séquence "StMichel"



HS



TVL1

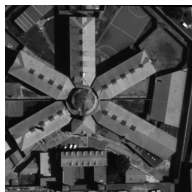


LK

	temps	erreur de reprojection		erreur moyen
V.T.	0.00 s	1.418	niveaux de gris	0.000 pixels
HS	1.16 s	0.945	niveaux de gris	0.286 pixels
TVL1	0.91 s	0.986	niveaux de gris	0.190 pixels
LK	0.29 s	1.250	niveaux de gris	0.331 pixels

# Comparaison des méthodes

Séquence "StMichel"



HS



TVL1

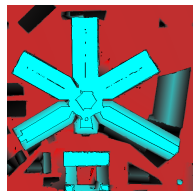


LK

	temps	erreur de reprojection	erreur moyen
V.T.	0.00 s	1.418 niveaux de gris	0.000 pixels
HS	1.16 s	0.945 niveaux de gris	0.286 pixels
TVL1	0.91 s	0.986 niveaux de gris	0.190 pixels
LK	0.29 s	1.250 niveaux de gris	0.331 pixels

# Comparaison des méthodes

Séquence "StMichel"



V.T.



HS



TVL1



LK

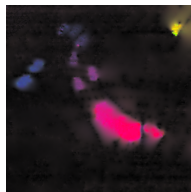
	temps	erreur de reprojection	erreur moyen
V.T.	0.00 s	1.418 niveaux de gris	0.000 pixels
HS	1.16 s	0.945 niveaux de gris	0.286 pixels
TVL1	0.91 s	0.986 niveaux de gris	0.190 pixels
LK	0.29 s	1.250 niveaux de gris	0.331 pixels

# Comparaison des méthodes

Séquence "Ettlinger"



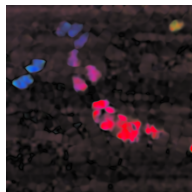
$I_1$



HS



TVL1



LK

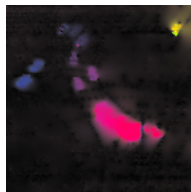
	temps	erreur de réprojection	erreur moyen
V.T.	-	-	-
HS	0.62 s	6.094 niveaux de gris	-
TVL1	1.50 s	6.154 niveaux de gris	-
LK	0.54 s	7.504 niveaux de gris	-

# Comparaison des méthodes

Séquence "Ettlinger"



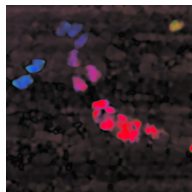
$I_2$



HS



TVL1



LK

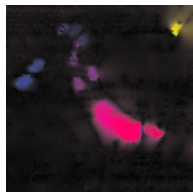
	temps	erreur de réprojection	erreur moyen
V.T.	-	-	-
HS	0.62 s	6.094 niveaux de gris	-
TVL1	1.50 s	6.154 niveaux de gris	-
LK	0.54 s	7.504 niveaux de gris	-

# Comparaison des méthodes

Séquence "Ettlinger"



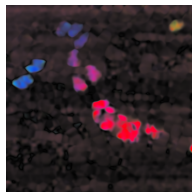
$I_2 - I_1$



HS



TVL1

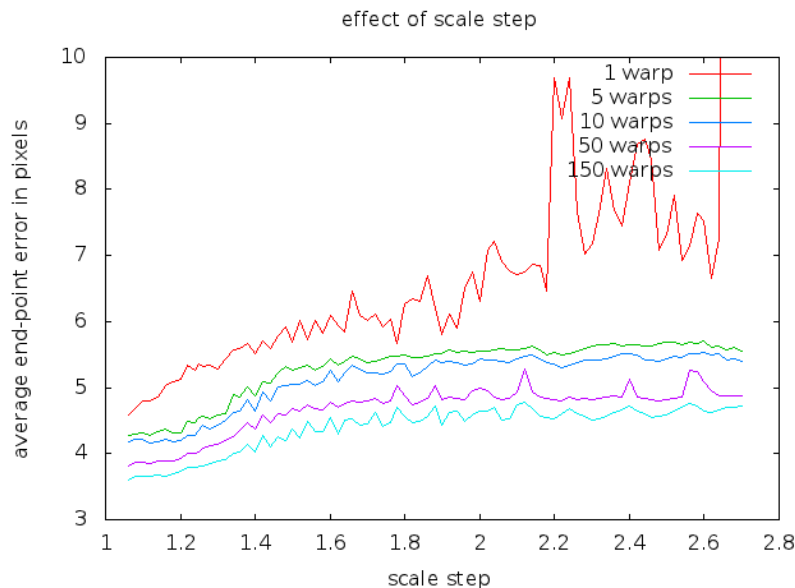


LK

	temps	erreur de reprojection	erreur moyen
V.T.	-	-	-
HS	0.62 s	6.094 niveaux de gris	-
TVL1	1.50 s	6.154 niveaux de gris	-
LK	0.54 s	7.504 niveaux de gris	-

# Estimation des paramètres par défaut

Recherche exhaustive





# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9



image originale 1/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9

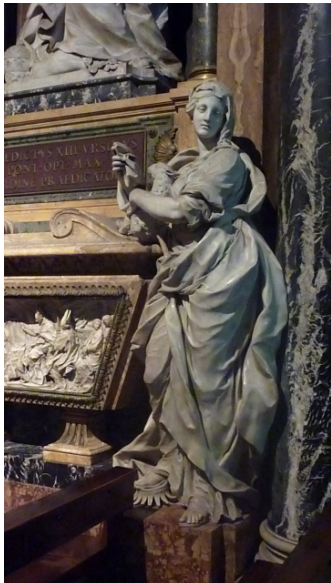


image originale 2/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9



image originale 3/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9



image originale 4/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9



image originale 5/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9



image originale 6/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9



image originale 7/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9



image originale 8/9



# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9



image originale 9/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9, temps HS=10x8=80s



image originale 1/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9, temps HS=10x8=80s

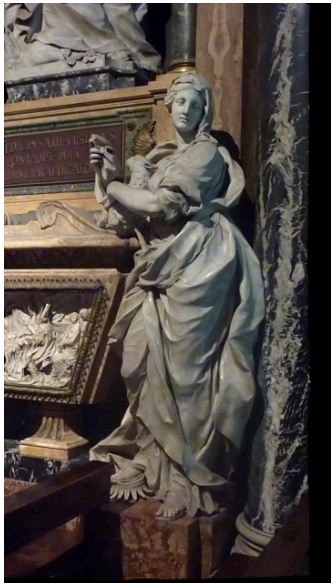


image enregistrée 2/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9, temps HS=10x8=80s



image enregistrée 3/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9, temps HS=10x8=80s

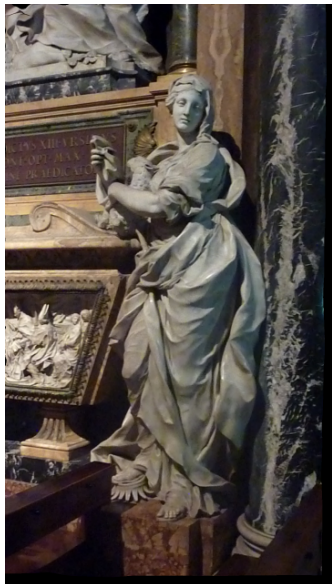


image enregistrée 4/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9, temps HS=10x8=80s



image enregistrée 5/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9, temps HS=10x8=80s



image enregistrée 6/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9, temps HS=10x8=80s



image enregistrée 7/9



# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9, temps HS=10x8=80s



image enregistrée 8/9

# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9, temps HS=10x8=80s

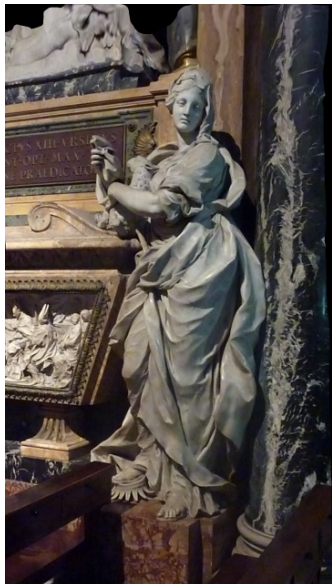


image enregistrée 9/9

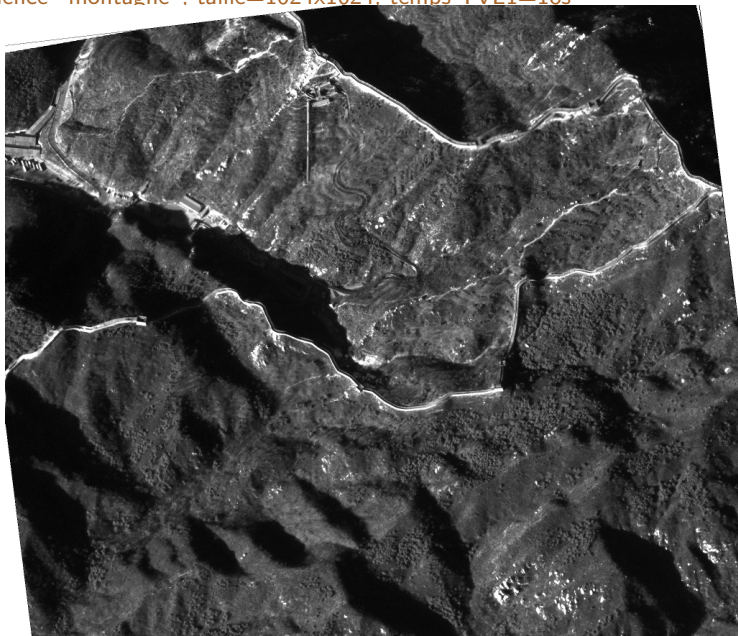
# Application : débruitage de rafales d'images

Séquence "statue", taille=1080x1920x9, temps HS=10x8=80s



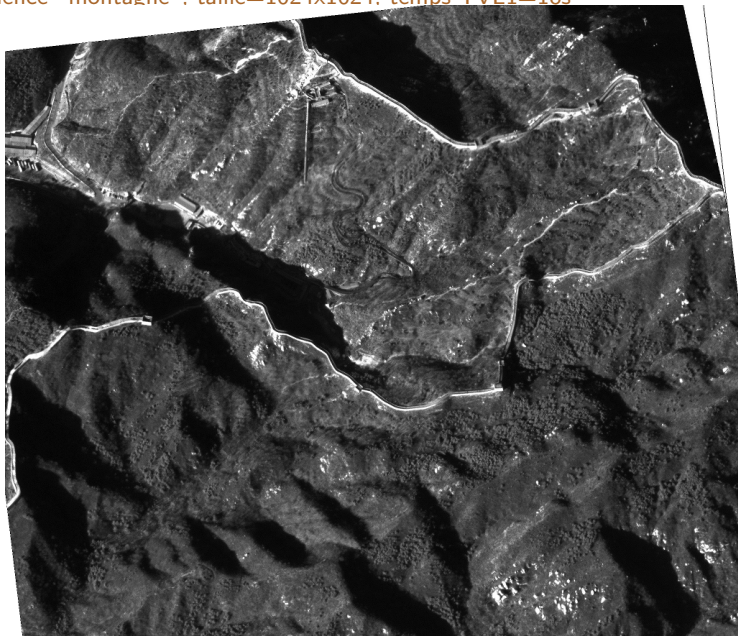
# Application : stéréo sans réctification

Séquence "montagne", taille=1024x1024, temps TVL1=18s



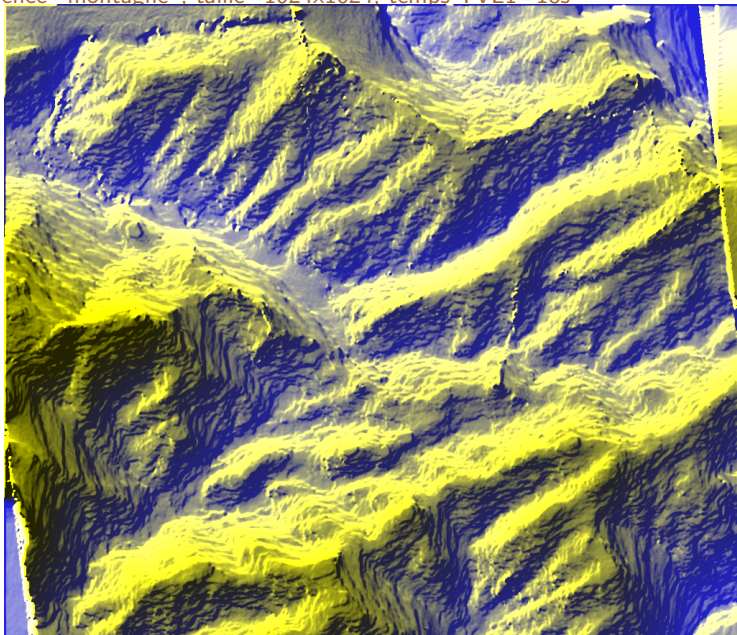
# Application : stéréo sans réctification

Séquence "montagne". taille=1024x1024. temps TVL1=18s



# Application : stéréo sans réctification

Séquence "montagne", taille=1024x1024, temps TVL1=18s



# Application : stéréo sans réctification

Séquence "ville", taille=1024x1024, temps TVL1=18s



# Application : stéréo sans réctification

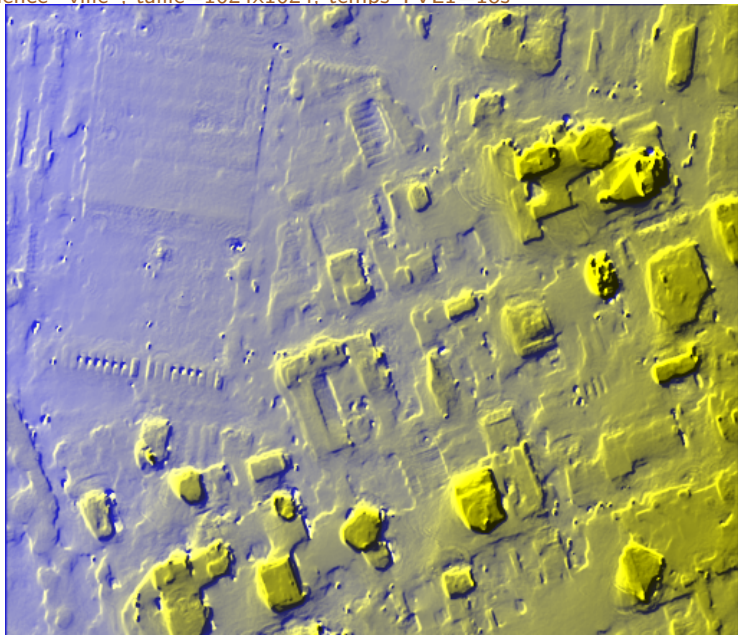
Séquence "ville", taille=1024x1024, temps TVL1=18s





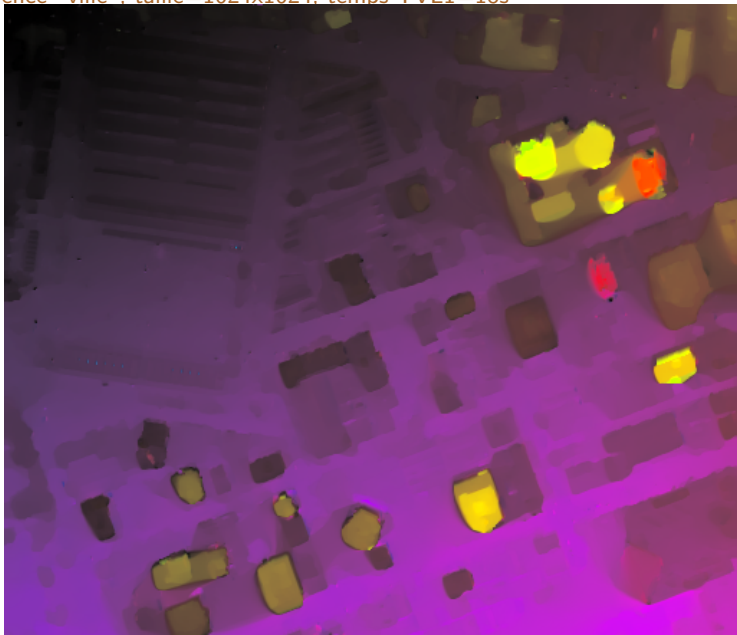
# Application : stéréo sans réctification

Séquence "ville", taille=1024x1024, temps TVL1=18s



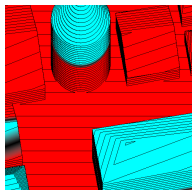
# Application : stéréo sans réctification

Séquence "ville", taille=1024x1024, temps TVL1=18s

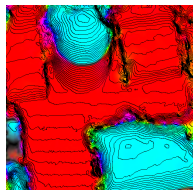


# Conclusion

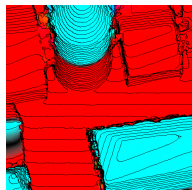
- Plusieurs méthodes avec des résultats similaires



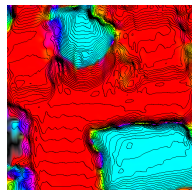
vérité terrain



HS : flou



TVL1 : discontinu

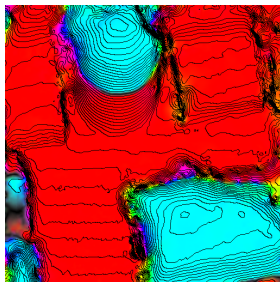


LK : local

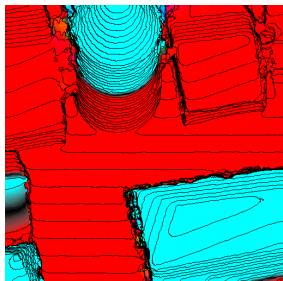
- Rapport qualité/temps d'exécution facile à gérer
- La pyramide multi-échelle est essentielle.



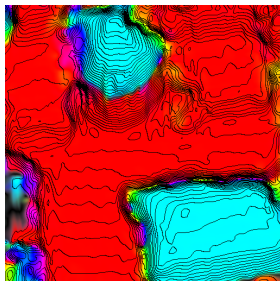
verité terrain



HS : flou



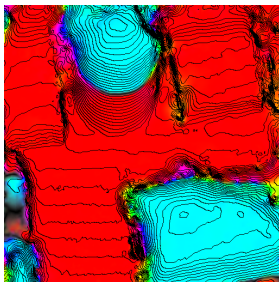
TVL1 : discontinu



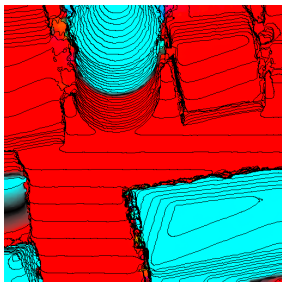
LK : local



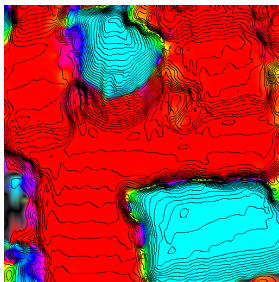
verité terrain



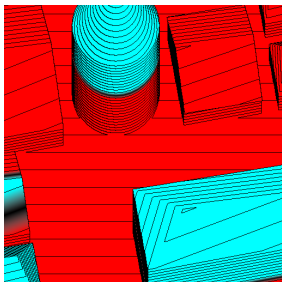
HS : flou



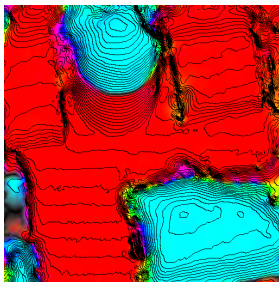
TVL1 : discontinu



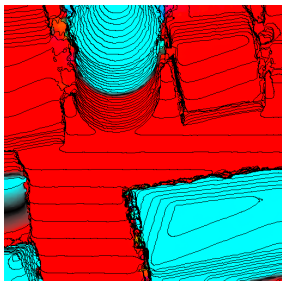
LK : local



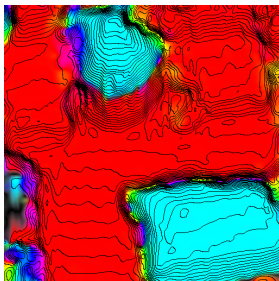
verité terrain



HS : flou



TVL1 : discontinu



LK : local